PATENT COOPERATION TREATY

From the INTERNATIONAL BUREAU

PCT

NOTIFICATION OF ELECTION

(PCT Rule 61.2)

To:

Assistant Commissioner for Patents
United States Patent and Trademark Office
Box PCT
Washington, D.C.20231
ETATS-UNIS D'AMERIQUE

in its capacity as elected Office

Date of mailing (day/month/year) 26 July 2000 (26.07.00)	
International application No. PCT/SE99/02169	**Applicant's or agent's file reference** 2008295
International filing date (day/month/year) 23 November 1999 (23.11.99)	**Priority date** (day/month/year) 24 November 1998 (24.11.98)
Applicant DANIELSON, Magnus et al	

1. The designated Office is hereby notified of its election made:

[X] in the demand filed with the International Preliminary Examining Authority on:

20 June 2000 (20.06.00)

[ ] in a notice effecting later election filed with the International Bureau on:

2. The election [X] was

[ ] was not

made before the expiration of 19 months from the priority date or, where Rule 32 applies, within the time limit under Rule 32.2(b).

The International Bureau of WIPO **34, chemin des Colombettes** **1211 Geneva 20, Switzerland**	Authorized officer F. Baechler
Facsimile No.: (41-22) 740.14.35	Telephone No.: (41-22) 338.83.38

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 :
H04L 12/43, 12/52, H04Q 11/04, H04L 12/56

A1

(11) International Publication Number: WO 00/31925

(43) International Publication Date: 2 June 2000 (02.06.00)

(21) International Application Number: PCT/SE99/02169

(22) International Filing Date: 23 November 1999 (23.11.99)

(30) Priority Data:
9804023-1 24 November 1998 (24.11.98) SE

(71) Applicant *(for all designated States except US)*: NET INSIGHT AB [SE/SE]; P.O. Box 42093, S-126 14 Stockholm (SE).

(72) Inventors; and

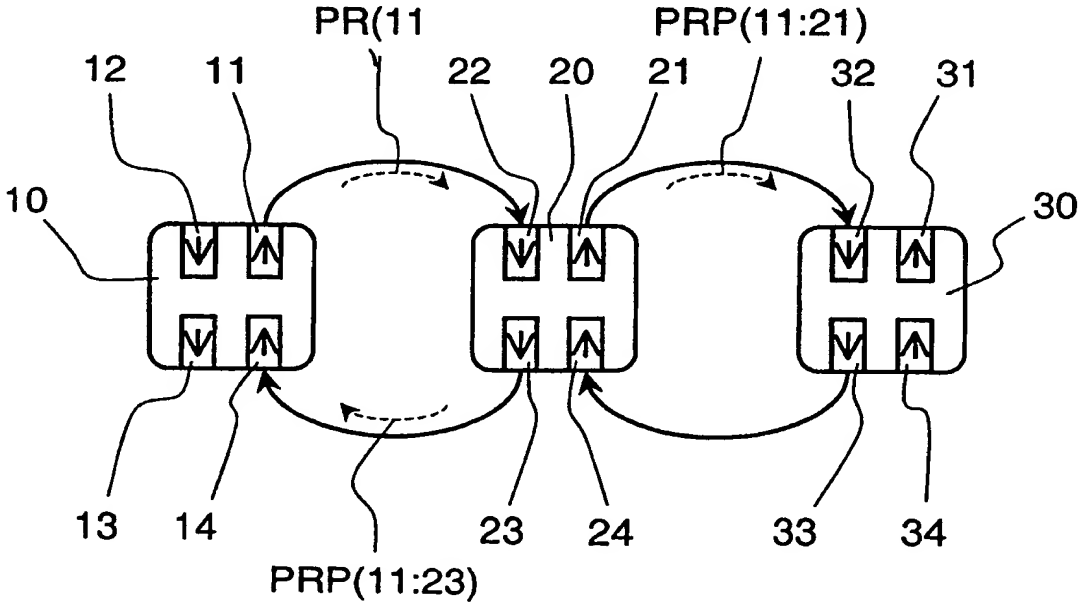
(75) Inventors/Applicants *(for US only)*: DANIELSON, Magnus [SE/SE]; Kyrkvägen 3 A, S-182 74 Stocksund (SE). HOLMLUND, Mattias [SE/SE]; Robert Almströmsgatan 6, S-113 36 Stockholm (SE). TESSIER, Stéphane [FR/SE]; Professorslingan 11, S-104 05 Stockholm (SE).

(74) Agent: AWAPATENT AB; P.O. Box 45086, S-104 30 Stockholm (SE).

(81) Designated States: AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), DM, EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHODS AND SYSTEMS FOR DETERMINING NETWORK TOPOLOGY



(57) Abstract

The present invention relates to a method and a system for determining the topology of a network of nodes that are interconnected via unidirectional connections. According to the invention, the existence of a network loop within said network is determined using message forwarding among said nodes, and, as a result, information related to the existence of said network loop is distributed to nodes within said network.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 99/02169

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: H04L 12/43, H04L 12/52, H04Q 11/04, H04L 12/56
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: H04L, H04Q, H04J

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GB 2133952 A (INTERNATIONAL STANDARD ELECTRIC CORPORATION), 1 August 1984 (01.08.84), page 1, line 36 - line 44; page 1, line 60 - page 2, line 6; page 2, line 62 - page 3, line 3, figures 1,2, claims 1-19, page 3, line 29 - line 33; page 4, line 19 - line 28	1-41
	--	
A	US 4287592 A (DANIEL J. PAULISH ET AL), 1 Sept 1981 (01.09.81), column 7, line 29 - line 36; column 8, line 19 - line 57, figures 3,4	1-41
	--	

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" erlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
17 April 2000	19 -04- 2000
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer Erik Johannesson/CL Telephone No. +46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 99/02169

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 5732086 A (SONG-CHYAU S. LIANG ET AL), 24 March 1998 (24.03.98), column 1, line 4 - line 30; column 2, line 51 - column 3, line 14, claims 1-10	1-41
A	US 4672373 A (KINJI MORI ET AL), 9 June 1987 (09.06.87), claims 1-11	1-41
A	US 5440540 A (WILHELM KREMER), 8 August 1995 (08.08.95), claims 1-36, abstract	1-41
A	WO 9727718 A1 (NEWBRIDGE NETWORKS CORPORATION), 31 July 1997 (31.07.97), see whole document	1-41
A	WO 9713344 A1 (TELEFONAKTIEBOLAGET LM ERICSSON), 10 April 1997 (10.04.97), see whole document	1-41

INTERNATIONAL SEARCH REPORT

Information on patent family members

02/12/99

International application No.
PCT/SE 99/02169

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
GB 2133952 A	01/08/84	AU 566427 B	22/10/87
		AU 2195783 A	28/06/84
		BE 895438 A	22/06/83
		US 4553233 A	12/11/85
US 4287592 A	01/09/81	NONE	
US 5732086 A	24/03/98	NONE	
US 4672373 A	09/06/87	CA 1237803 A	07/06/88
		DE 3486204 D,T	23/12/93
		DE 3486456 D,T	30/04/98
		EP 0147789 A,B	10/07/85
		EP 0502555 A,B	09/09/92
		JP 2036218 C	28/03/96
		JP 7067112 B	19/07/95
		JP 60134647 A	17/07/85
US 5440540 A	08/08/95	CA 2065463 A,C	27/09/93
WO 9727718 A1	31/07/97	AU 1433697 A	20/08/97
		AU 6466296 A	10/02/97
		CA 2244073 A	31/07/97
		EP 0837835 A	29/04/98
		EP 0875123 A	04/11/98
		GB 9601692 D	00/00/00
		JP 11508869 T	03/08/99
WO 9713344 A1	10/04/97	AU 7079796 A	28/04/97
		AU PN573795 D	00/00/00
		CA 2231380 A	10/04/97
		EP 0853850 A	22/07/98

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04Q 11/04, H04L 12/43 — A1

(11) International Publication Number: WO 97/27718

(43) International Publication Date: 31 July 1997 (31.07.97)

(21) International Application Number: PCT/CA97/00052

(22) International Filing Date: 27 January 1997 (27.01.97)

(30) Priority Data:
9601692.8 27 January 1996 (27.01.96) GB

(71) Applicant *(for all designated States except US)*: NEWBRIDGE NETWORKS CORPORATION [CA/CA]; 600 March Road, P.O. Box 13600, Kanata, Ontario K2K 2E6 (CA).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: MITCHELL, Charles [CA/CA]; 78 Summerwalk Place, Nepean, Ontario K2G 5Y4 (CA).

(74) Agent: MITCHELL, Richard, J.; Marks & Clerk, Station B, P.O. Box 957, Ottawa, Ontario K1P 5S7 (CA).

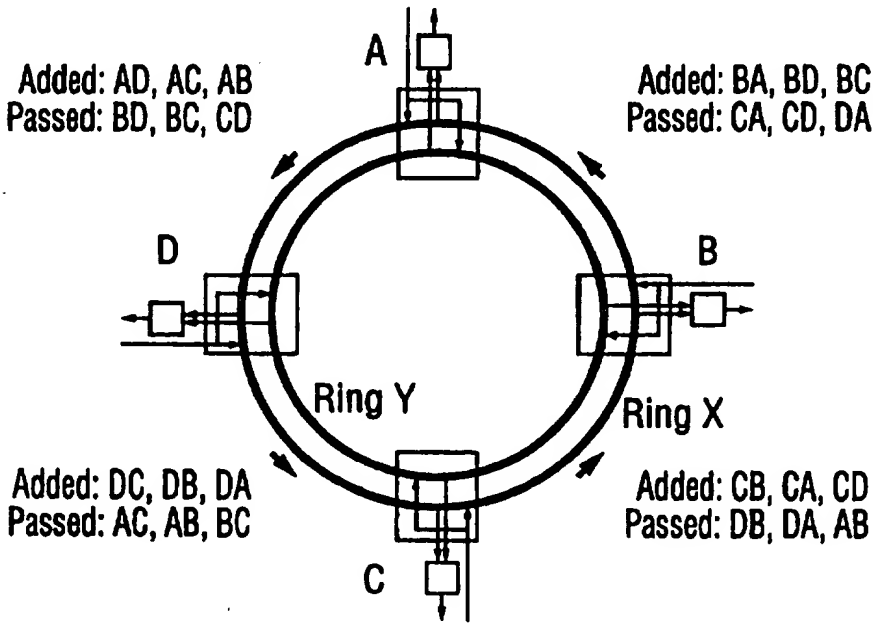
(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ATM NETWORK WITH RING ARCHITECTURE



(57) Abstract

A communications network comprises a plurality of nodes interconnected by a pair of packet-based rings, for example ATM rings. At each said node there is a ring interface comprising a device for extracting packets from either ring destined for that node, a device for adding packets from that node to either ring, and a device for passing directly through packets on each ring not destined for that node. This arrangement establishes a virtual mesh structure with offers good fault recovery without unnecessary wastage of bandwidth.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

ATM NETWORK WITH RING ARCHITECTURE

This invention relates to packet-based networks, such as ATM (Asynchronous Transfer Mode) networks, having a ring architecture. The invention is applicable to both local and wide area networks.

There is a need within networks to ensure that network operation is not interrupted in the event of infrastructure failures, such as cable cuts and/or equipment failure. The current state of the art is to provide protection with SONET Ring architectures. Bellcore standard GR-1230-CORE describes a Bi-direction Line Switched Line Switched Ring and GR-1400-CORE describes a Uni-directional Path Switched Ring.

In the case of ATM networks, the current state of the art is to send ATM payloads over SONET ring structures. As this is an expensive solution, there is a need to provide a simpler mechanism for ATM network protection. While it is possible to map SONET Ring structures directly into an ATM implementation using the ATM Switch to perform the necessary bridging and re-configuration, the result is an inefficient use of the ATM Switch and a slow response time.

An object of the invention is to provide a wide area network architecture that alleviates the afore-mentioned problems of the prior art.

According to the present invention there is provided a communications network a communications network comprising a plurality of nodes interconnected in a ring architecture. A pair of counter-rotating packet-based rings interconnect the nodes to establish virtual connections therebetween. Each node includes a ring interface and a packet switch connected thereto. The ring interface comprises means for extracting packets from either ring destined for the packet switch, means for adding outgoing packets the packet switch to either ring, and means for passing directly through packets on each ring not destined for the node. This results in a virtual mesh structure being established between said nodes.

The rings can either be unidirectional or bi-directional. Of course, it will be understood by one skilled in the art that additional rings can be provided without detracting from the advantages offered by the invention.

The described ring structure allows bandwidth to be efficiently shared between the nodes on the ring without the wastage inherent in a hard-wired mesh and without using up valuable ATM switch resources since onward cells are passed directly through the interface without being forwarded to the ATM switch.

The invention also provides a method of establishing communication over a network wherein a plurality of nodes are interconnected in a ring architecture. A pair of counter-rotating packet-based rings are established to interconnect the nodes and provide virtual connections therebetween At each node packets destined for the node are extracted from the rings, outgoing packets from the node are added to the rings, and packets on each ring not destined for that node are passed directly through. This results a virtual mesh structure being established between the nodes.

The invention will now be described in more detail, by way of example only, with reference to the accompanying drawings, in which:-

Figure 1 shows a four node mesh network;

Figure 2 shows a four node ring network;

Figure 3 shows a ring interface using OC-12c Interface Cards for a Newbridge Networks 36170 ATM switch;

Figure 4 shows a simple uni-directional ATM Switched Ring (UASR);

Figure 5 shows an example of a double ring fault;

Figure 6 shows another example of a double ring fault;

Figure 7a shows a bi-directional ATM Switched Ring (BASR);

Figure 7b shows a portion of the ring shown in figure 7a;

Figure 8 shows a SONET-ATM ring overlay.

Figure 9 is a block diagram of an OC-12c card; and

Figure 1 shows a four-node mesh network forming a SONET ring. There are six possible bi-directional connections between the nodes. i.e. A-B, A-C, A-D, B-C, B-D and C-D. In a mesh network such as this, six point-to-point links are required to complete the

mesh. If these links are OC-3c interfaces with approximately 150 Mbps capacity, since the bandwidth of these individual links is fixed, no advantage can be taken of the variable bandwidth demands of the nodes. If node A on average requires 50 Mbps to node B, 100 Mbps to node C and only 20 Mpbs to node D, the unused capacity to these links is wasted.

While it is possible to route traffic via alternative routes to try and make use of this bandwidth, this is a software intense task that requires extra CAC (Connection Admission Control) calculations and uses up valuable switch resources at these tandem nodes.

In Figure 2, the four nodes A, B, C, D are interconnected by an OC-12c ATM Ring. An OC-12c interface has four times the capacity of a OC-3c link, i.e. approx. 600 Mbps. In this case, each segment of the ring carries six virtual paths, resulting in a logical mesh, i.e. each node has a direct virtual path to every other node. The advantage of this arrangement is that all of the paths are carried on the one link. This means that the nodes using less bandwidth free up the capacity for other nodes to use, and when required nodes can use up to the full bandwidth of the Ring, i.e. node A can send a full 600 Mbps to node C.

Two types of ATM ring can be used: a) A Uni-directional ATM Switched Ring (UASR), and b) A Bi-directional ATM Switched Ring (BASR).

In the UASR, the same traffic is transmitted in opposite directions on two counter-rotating Rings. At each node ingress ATM cells come in on a standard cell relay interface. Cells destined for that node are "dropped" into the switching fabric of the node while the remaining cells are queued with the cells to be "added" by that node and sent out the egress port of the standard cell relay interface.

Figure 3 shows how this ring interface maps onto a Newbridge Networks Corporation 36170 OC-12c card at each node. Rings X and Y pass through OC-12c cards 1 and 2 respectively. These are connected through a Fabric Interface Card (FIC) 3 to a Newbridge Networks Corporation 36170 ATM switching Core 4.

On the ingress side of each ring, the Ingress Cell Controller (ICC) 5 on the OC-12c card determines whether the incoming cells are to be dropped onto the UP_ISL by looking at the cell's VPI field. Each node on the ring is addressed with a range of VPI values. The Fabric Interface Card 3 receives drop traffic from each Ring and has the ability to select cells from either, or both Rings.

This ability is provided by, for example, Newbridge Networks Corporation Diablo ASIC 6, which is a two-input to one output cell switching engine. The inputs of the Diablo 6 are connected to the respective rings X,Y. The "through" cells are queued with cells to be added from the DOWN_ISL (Inter-Shelf Link), again using the Diablo switching ASIC 6 and Egress cell controller 7. Alternatively, all traffic from the ring can be brought up the UP_ISL to the switching core where the cells to be dropped are sent down to their appropriate ISL, and through cells are returned down the OC-12 cards DOWN_ISL along with cells to be added to the ring. In either case, an important feature is that traffic to be dropped at a node can be selected from either or both rings.

The rules for receiving data off the ring are quite different from a time division multiplex ring, such as a SONET ring, for example.

Figure 4 shows a simple UASR. If both the X and Y rings are fault-free, any node on the ring is free to receive data from either ring, i.e. the Diablo switch 6 on the FIC 3 selects only one of its inputs to switch directly to its output. If a fault occurs on one of the rings, then all nodes program the Diablo 6 to receive data off the other fault-free ring. If both rings have faults, then all nodes receive data from both rings by enabling both inputs to the Diablo 4.

The common cases are where both rings have faults, and the rings can recover fully, a cable cut on the span connecting two nodes, and a node going out of service. Other double faults can leave the ring operating in a diminished capacity.

It will be appreciated that with ATM data transmission, cells are only transmitted when there is data to send. Consequently, as long as the identical cells are not being received on the two rings, any ATM cell destined for a particular node can be received. In order to ensure that identical cells are not received at any node, it is important to ensure

that when both rings have faults, any node that detects a fault on its incoming link squelch its corresponding counter-rotating link on the same span.

There are several cases of faults on both rings that need to be considered. The simplest and most common case is that of a cable cut in a span connecting two nodes in the Ring. Figure 5 shows illustrates this case. In this case the nodes A, D surrounding the cut will detect the cut on the incoming ring and unnecessarily squelch the traffic in the counter rotating ring. Unnecessarily, in this case because the cell flow in the counter rotating direction could not have been received by any other node due to the cut.

A more generalized case is shown in Figure 6. In this case a double fault results in segregated islands of links. Each island is capable of full-duplex communication with other members in the island. For example in Figure 6, nodes A through F will enjoy complete communication with each other. Note that if nodes G and H had not squelched their traffic, cells from G and H would be received twice at nodes A through F (in the double Ring fault case, all nodes receive cells from both Rings).

During normal operation, in-band ATM cells or part the SONET can be used to indicate to all nodes on the ring that the ring is fault. In the latter case, any node that has detected a fault on one of its incoming links sets an indication to the next node on the uni-directional ring. Any node receiving such and indication on its incoming link sets an indication to the next node on the uni-directional ring.

The bi-directional ATM Switched Ring (BASR) shown in Figures 7a and 7b is a further refinement of the ATM Ring concept where a small change in approach allows for the use of the "extra" bandwidth on the Ring. Extra bandwidth is freed up on the ring by not broadcasting the same cells on both Rings, but selectively sending Add cells on one of the two Rings. Figures 7a and 7b show how node A can transmit cells on the X Ring to node D and transmit cells on the Y Ring to nodes B and C. As a minimum, this results in less traffic on the rings and therefore less congestion at the queuing points. For this type of ring the receiving nodes are programmed to always receive cells from both rings. In the event of ring failures, it is the transmitting function that alters its behavior. Any ring failure will cause all nodes to change from transmitting selectively on a particular ring, to broadcasting all cells on both rings.

The examples discussed so far have been for OC-12c Rings. OC-12c rings map well into the Newbridge Networks Corporation's 36170 ISL (Inter-Shelf Link) architecture. The particularly attractive feature is that the through traffic on the rings does not go into the switching core and that it takes only one ISL to support an OC-12c Ring.

There are two fundamentally different applications for OC-12c Rings. The first is the "backbone" application where OC-12c Rings are used to connect large switches to form a network backbone. The other is a "collector" application where many smaller nodes feed the ring with smaller bandwidths.

In the backbone case, the number of switches on a ring would normally be limited to four or five nodes. This limits the delay and CDV and it also justifies the size of the add and drop rates, i.e. ISLs (smaller add/drop rates would fall under the collector application).

Additional capacity is readily achieved by adding parallel rings. If the number of nodes on the backbone begins to exceed the optimum number, the parallel rings would be staggered to pick up different nodes.

The collector application by definition requires less add/drop bandwidth and generally picks up traffic from a larger number of smaller nodes. In this case the add/drop bandwidth need not be a full ISL. An OC-12c ring card can be made to fit in a UCS slot. This would allow even Standalone 36170 Shelves to be connect in a ring configuration. An application here may be pick up low bandwidth Frame Relay traffic onto a feeder ring (which would connect to a larger ATM backbone ring).

Transmitting selectively out the two OC-12c cards can be accomplished by:

1) a combination of programming the headers appropriately at all Ingress cards and programming the point-to-point filters to select one or both of the OC-12c card addresses, or

2) a hardware ECC function that identifies and enables/disables groups of connections. In this case Ingress connection programming would not change.

ATM Rings can also be made to work well using OC-3c interface cards. In this case, through traffic on the Ring would have to go through the switching core.

Transmitting out both OC-3c cards is readily accomplished by programming the point-to-point filters. Multicast is also easily handled by programming the same multicast groups.

Ingress selection of which Ring to listen to is accomplished by programming the Diablo switches 6 on the FICs 3 to select the appropriate Add bus (or both in the case of a Ring failure).

Transmitting selectively out the two OC-3c cards is accomplished by a combination of programming the Newbridge headers appropriately at all Ingress cards and programming the point-to-point filters to select one or both of the OC-3c card addresses.

Figure 9 is a block more detailed diagram of an OC-12c card (1, 2) (Figure 3). Cells coming off the ring pass through PHY unit 10 to ICC unit 5. Cells not destined for the node pass through diablo unit 1, FIFO 12 and egress PHY unit 13. Cells destined for the node are extracted by Diablo 3 and passed to UP_ISL link 15.

Cells from the node destined for the ring passes through stealth unit 16 to diablo unit 2. Egress cell controller 7 controls the adding of cells from the node to those passing through.

Given that there exists a significant installed base of SONET rings, it is also important that the ATM ring work well with existing SONET Rings.

Figure 8 shows how an ATM ring can be overlaid on a SONET ring. This is important because ATM interfaces become increasingly difficult to design at the higher SONET rates. Interfaces at OC-12c are feasible and cost effective; however cell processing at the next step common step in the hierarchy, OC-48c, becomes extremely difficult to deal with. IN Figure 8, For example, an OC-12cATM ring is overlaid on an OC-48 SONET ring. BSS is an ATM ring with ATM VP functionality as per Bellcore GR-2837-CORE.

There will be demands to handle aggregate bandwidth of much greater than that offered by OC-12c interfaces. This can be achieved by running multiple OC-12c ATM rings overlaid on simple OC-48 or OC-192 SONET rings.

The described architecture provides good fault protection while minimizing wasted available bandwidth.

Although preferred embodiments of the invention have been illustrated in the accompanying drawings and described in the foregoing detailed description, it will be understood that the invention is not limited to the embodiments discloses, but is capable of numerous rearrangements, modifications and substitutions of parts and elements without departing from the spirit of the invention which is defined in the following claims.

Claims:

1. A communications network comprising a plurality of nodes interconnected in a ring architecture, characterized in that a pair of counter-rotating packet-based rings interconnect said nodes to establish virtual connections therebetween, and each said node includes a ring interface and a packet switch connected thereto, said ring interface comprising means for extracting packets from either ring destined therefor and passing them to the packet switch, means for adding outgoing packets from the packet switch to either ring. and means for passing packets directly through on each ring not destined for that node. whereby a virtual mesh structure is established between said nodes.

2. A network as claimed in claim 1, characterized in that said rings are unidirectional and counter-rotating.

3. A network as claimed in claim 1, characterized in that said rings are bi-directional.

4. A network as claimed in claim 3, wherein during normal operation said adding means selectively transmits packets on only one of said bi-directional rings, and in the event of a fault broadcasts packets both said rings.

5. A network as claimed in any of claims 1 to 3 characterized in that said network is an ATM network and said packets are ATM cells.

6. A network as claimed in any of claims 1 to 4 characterized in that said ATM network is overlaid onto a time division multiplexed (TDM) network.

7. A network as claimed in claim 5 wherein said TDM network is a SONET network.

8. A method of establishing communication over a network wherein a plurality of nodes are interconnected in a ring architecture, characterized in that a pair of counter-rotating packet-based rings are established to interconnect said nodes and provide virtual connections therebetween, and that at each node packets destined for the node are extracted from the rings, outgoing packets from the node are added to the rings, and packets on each ring not destined for that node are passed directly through. whereby a virtual mesh structure is established between said nodes.

9. A method as claimed in claim 8, characterized in that said rings are unidirectional and counter-rotating.

10. A method as claimed in claim 9, wherein when both said rings are fault–free said interface selects one of said rings to receive data, when one of said rings has a fault said interface device receives data off the other ring, and when both said rings have faults said interface device receives data of both said rings.

11. A method as claimed in claim 9 or 10 characterized in that upon detection of a fault, said interface device inhibits outgoing data transmission over a span to an adjacent node on which a fault has been detected.

12. A method as claimed in claim 8, characterized in that said rings are bi-directional.

13. A method as claimed in claim 11, wherein during normal operation packets are selectively transmitted on only one of said bi-directional rings, and in the event of a fault packets are broadcast on both said rings.

14. A method as claimed in any of claims 8 to 13, characterized in that said network is an ATM network and said packets are ATM cells.

15. A method as claimed in any of claims 8 to 14 characterized in that said ATM network is overlaid onto a time division multiplexed (TDM) network.

16. A method as claimed in claim 15 characterized in that said TDM network is a SONET network.

17. A method as claimed in 14, characterized in that fault information is carried by in-band ATM cells transmitted on the rings.

18. A method as claimed in 17, characterized in that fault information is carried by SONET overhead.

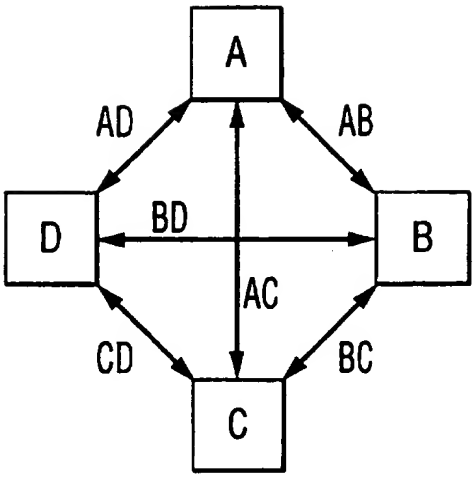

A
AD
AB
D
BD
B
AC
CD
BC
C

FIG. 1

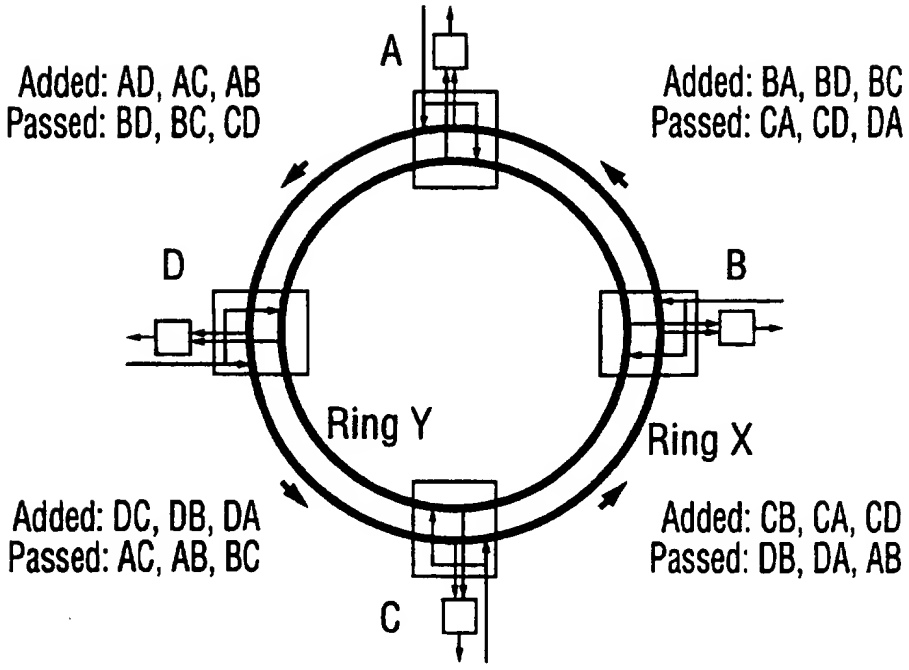
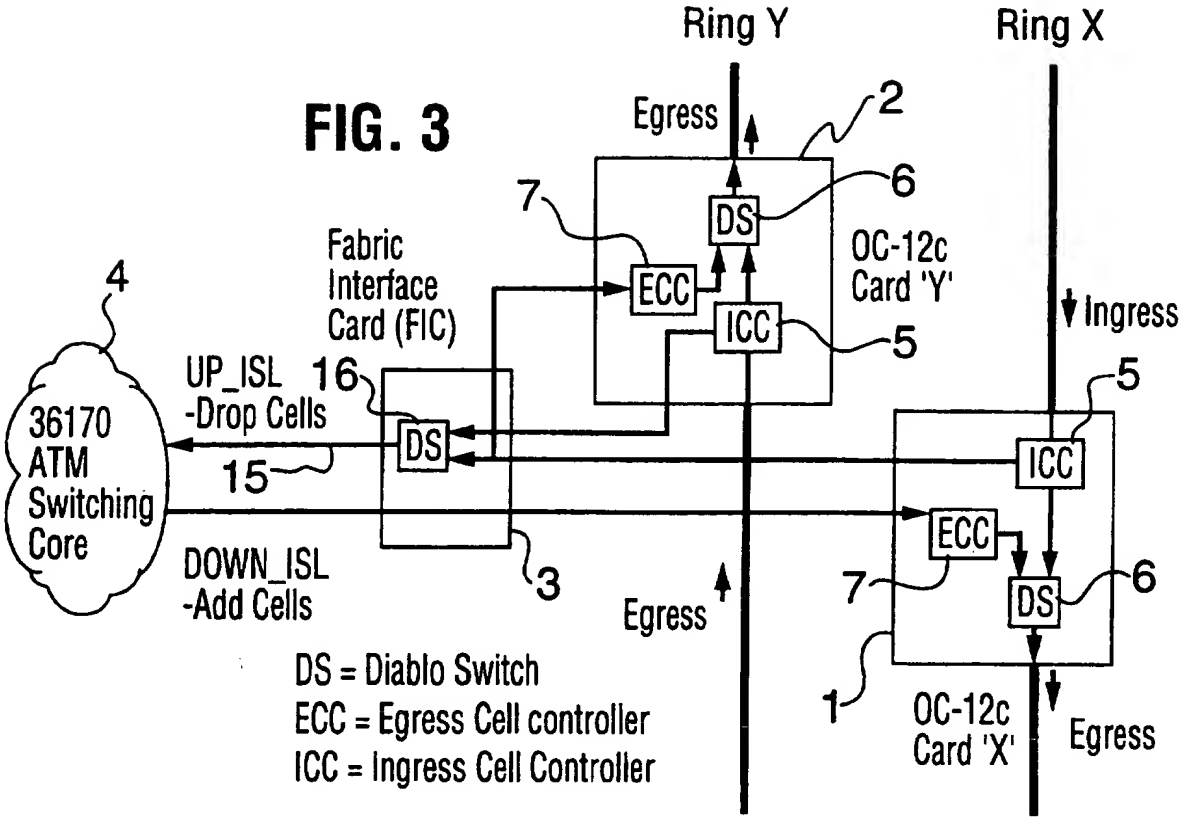

A
Added: AD, AC, AB
Passed: BD, BC, CD
Added: BA, BD, BC
Passed: CA, CD, DA
D
B
Ring Y
Ring X
Added: DC, DB, DA
Passed: AC, AB, BC
Added: CB, CA, CD
Passed: DB, DA, AB
C

FIG. 2

FIG. 3



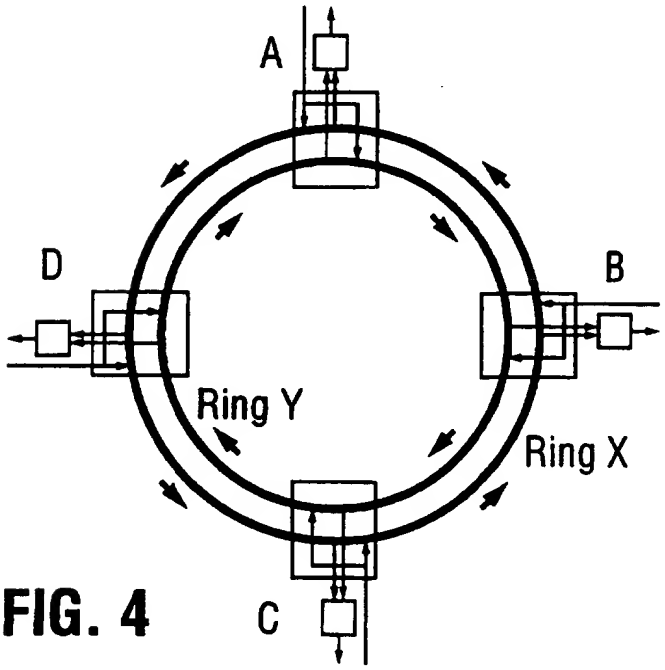


FIG. 4

Simple UASR:
1) TX all traffic on both Rings
2) If no failure
Rx all traffic from either
If one Ring fails
RX traffic from other Ring
If both Rings fail*
RX traffic from both Rings

*both directions of a span

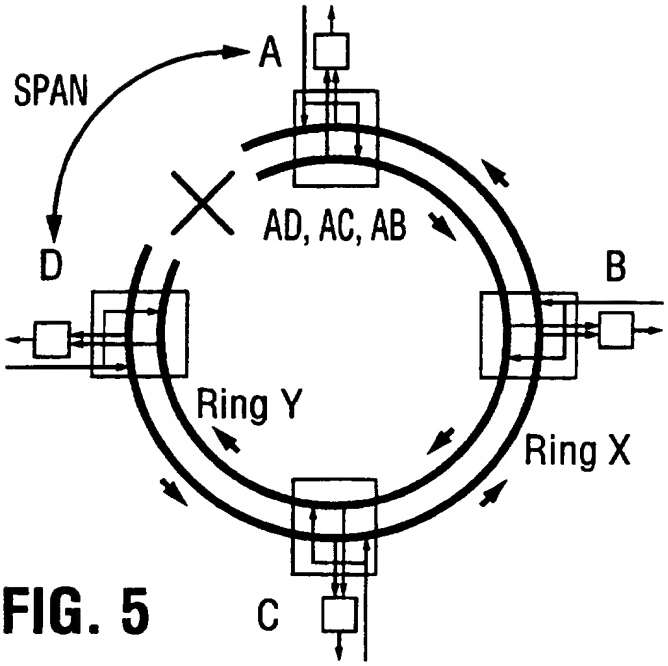

SPAN
A
AD, AC, AB
D
B
Ring Y
Ring X
C

FIG. 5

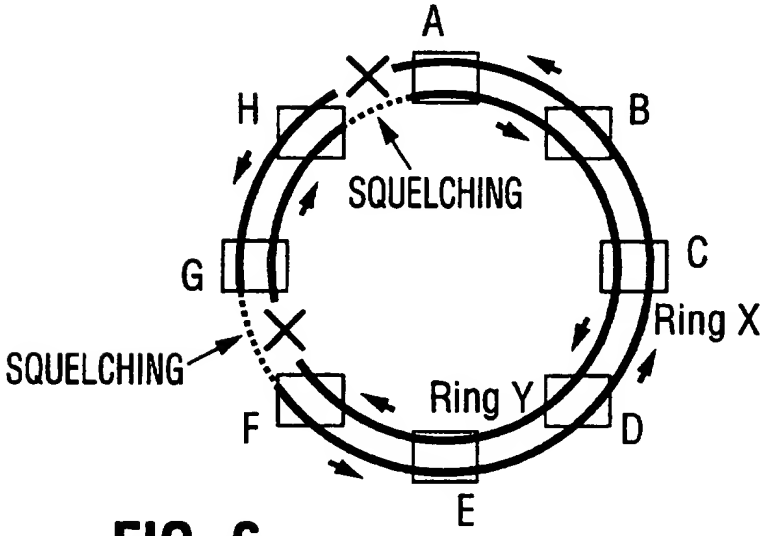

A
H
B
SQUELCHING
G
C
Ring X
SQUELCHING
Ring Y
F
D
E

FIG. 6

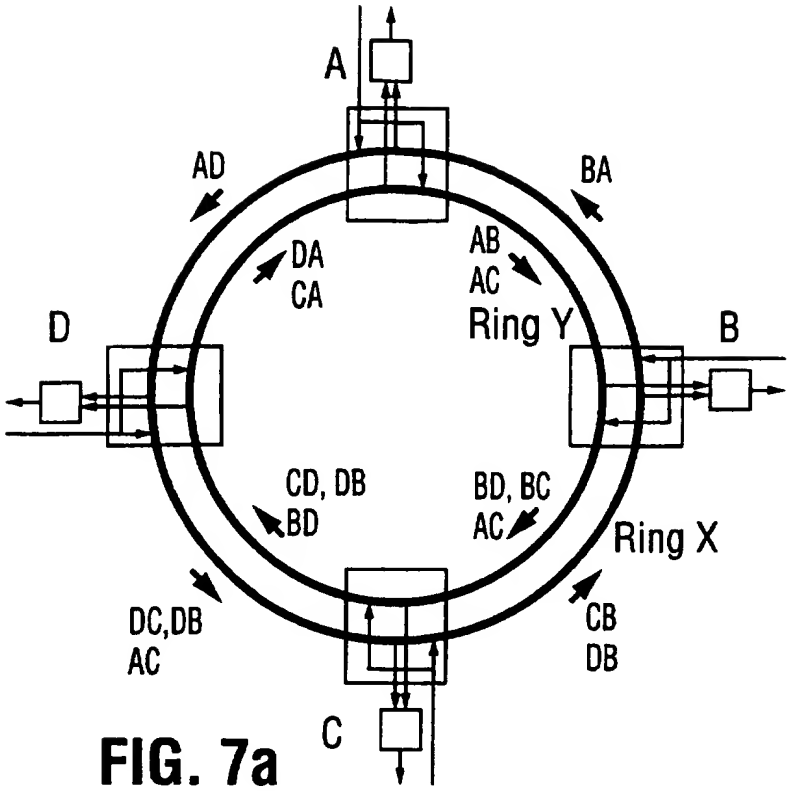


FIG. 7a

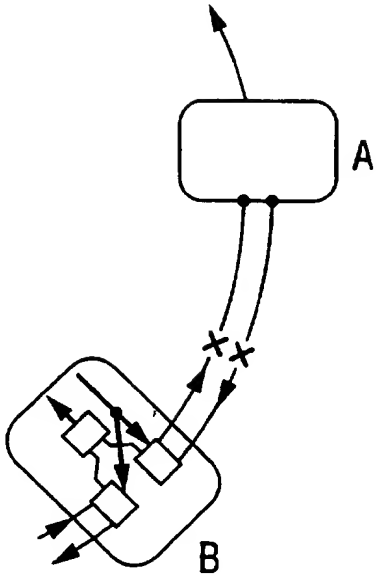


FIG. 7b

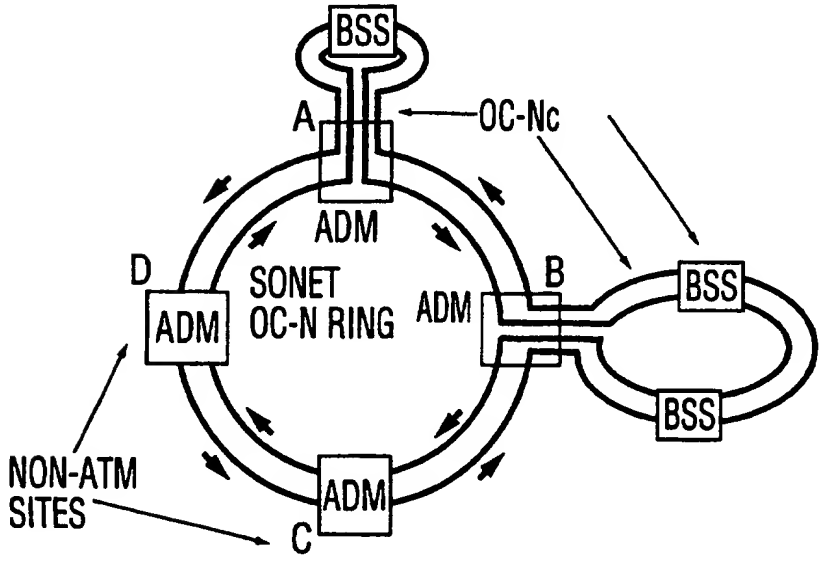


FIG. 8

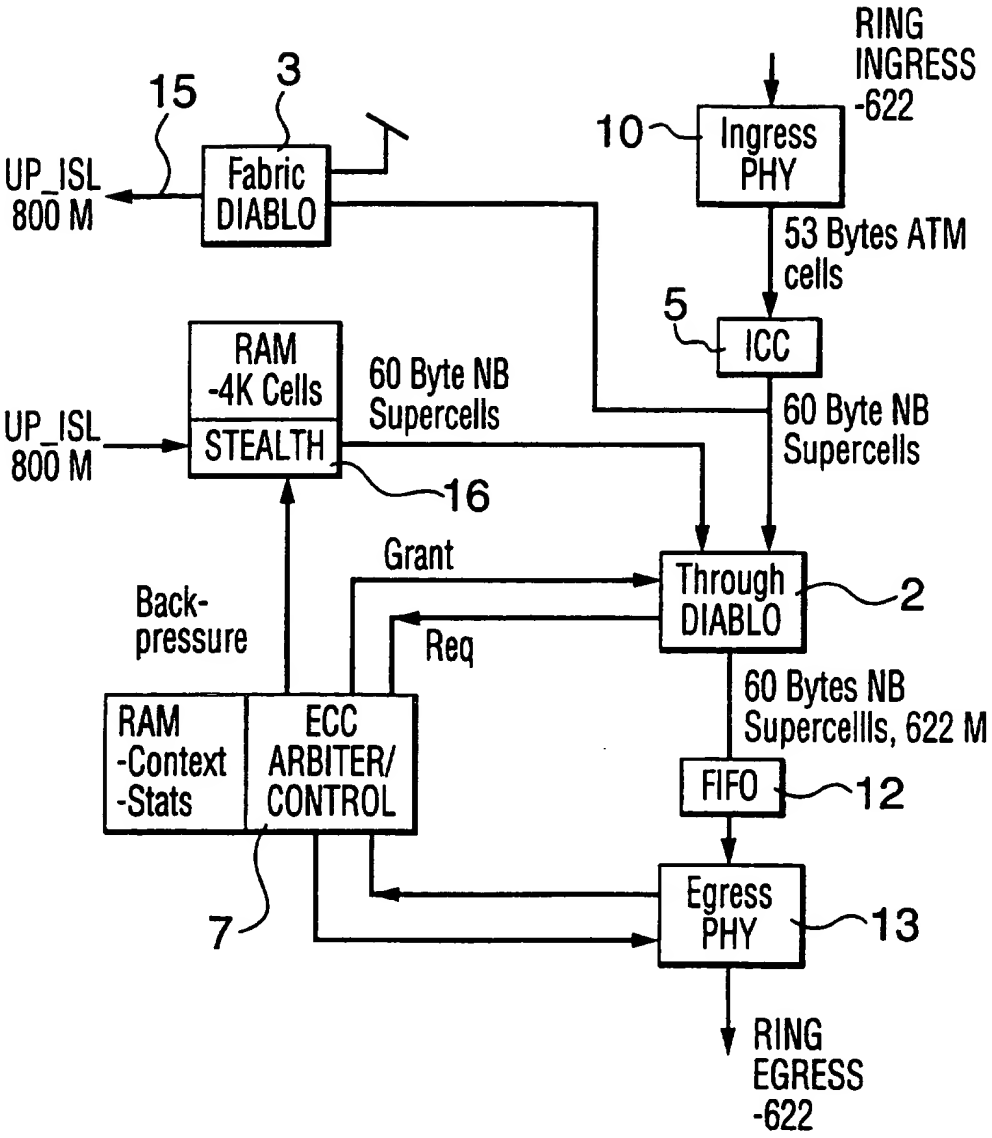

RING
INGRESS
-622
15
3
10
Ingress
PHY
UP_ISL
800 M
Fabric
DIABLO
53 Bytes ATM
cells
5
ICC
RAM
-4K Cells
60 Byte NB
Supercells
UP_ISL
800 M
STEALTH
60 Byte NB
Supercells
16
Grant
Through
DIABLO
2
Back-
pressure
Req
60 Bytes NB
Supercells, 622 M
RAM
-Context
-Stats
ECC
ARBITER/
CONTROL
FIFO
12
Egress
PHY
13
7
RING
EGRESS
-622

FIG. 9

INTERNATIONAL SEARCH REPORT

Inter nal Application No
CA 97/00052

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 H04Q11/04 H04L12/43

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 H04Q H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	COMMUNICATION FOR GLOBAL USERS, INCLUDING A COMMUNICATIONS THEORY MINI CONFERENCE ORLANDO, DEC. 6 - 9, 1992, vol. 1 - 2 - 03, 6 December 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 834-840, XP000357681 TSONG-HO WU ET AL: "A BROADBAND VIRTUAL PATH SONET/ATM SELF-HEALING RING ARCHITECTURE AND ITS ECONOMIC FEASIBILITY STUDY" *section 2* see page 837, column 1, line 7 - line 45 * section 3.1 * --- -/--	1-3,5-9, 12,14,15

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

3

Date of the actual completion of the international search	Date of mailing of the international search report
21 May 1997	09. 06. 97
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Authorized officer Gregori, S

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	COMMUNICATION FOR GLOBAL USERS, INCLUDING A COMMUNICATIONS THEORY MINI CONFERENCE ORLANDO, DEC. 6 - 9, 1992, vol. 1 - 2 - 03, 6 December 1992, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, pages 639-643, XP000357855 YOSHIO KAJIYAMA ET AL: "ATM SELF-HEALING RING" see page 639, column 1, line 19 - line 25 see page 639, column 2, line 7 - page 640, column 2, line 3 see page 640, column 2, line 29 - page 642, column 1, line 15 ---	1,3-8, 14-18
X	IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, vol. 12, no. 1, 1 January 1994, pages 171-177, XP000493898 YOSHIO KAJIYAMA ET AL: "AN ATM VP-BASED SELF-HEALING RING" see page 171, column 1, line 18 - column 2, line 16 see page 172, column 2, line 25 - line 32 see page 174, column 2, line 15 - page 175, column 2, line 14 -----	1,8-11, 13

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : H04L 12/42, 12/437, G06F 13/40, 15/173 — A1

(11) International Publication Number: **WO 97/13344**

(43) International Publication Date: 10 April 1997 (10.04.97)

(21) International Application Number: PCT/AU96/00621

(22) International Filing Date: 2 October 1996 (02.10.96)

(30) Priority Data:
PN 5737 2 October 1995 (02.10.95) AU

(71) Applicant *(for all designated States except US)*: TELEFON-AKTIEBOLAGET LM ERICSSON [SE/SE]; Telefonplan, S-126 25 Stockholm (SE).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: YIM, Ting, Shing [AU/AU]; 27 Tangemere Avenue, Tullamarine, VIC 3043 (AU).

(74) Agent: CARTER SMITH & BEADLE; Qantas House, 2 Railway Parade, Camberwell, VIC 3124 (AU).

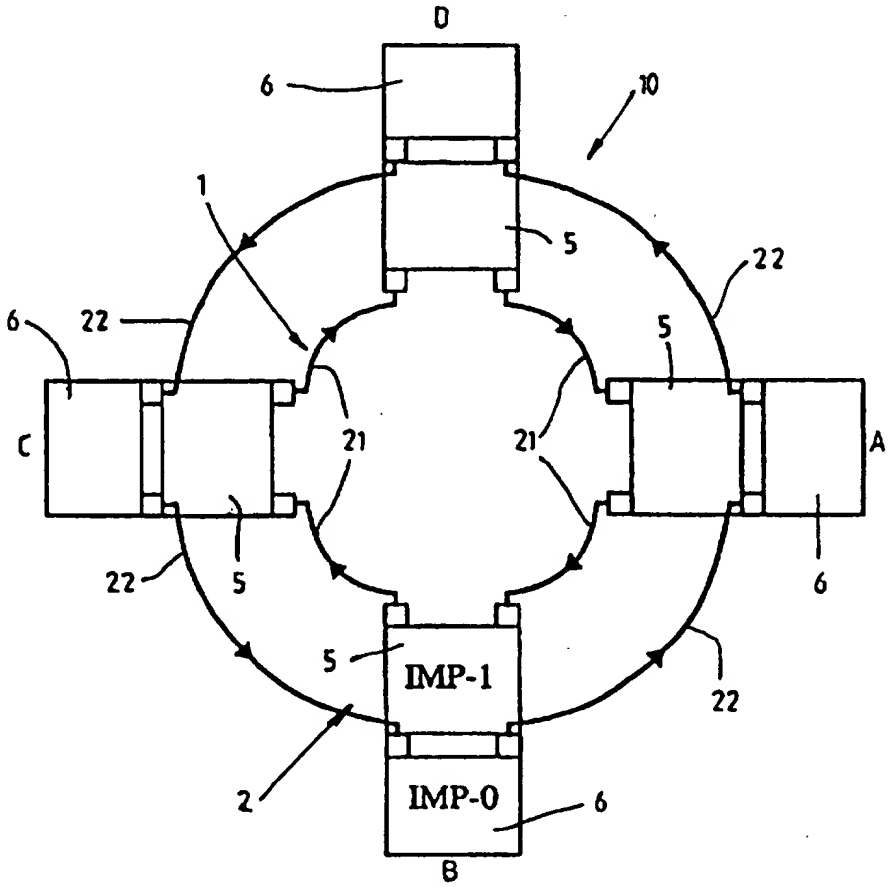
(81) Designated States: AU, CA, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: TRANSMITTING DATA BETWEEN MULTIPLE COMPUTER PROCESSORS

(57) Abstract

A communications system and method is provided in which data is transmitted between a plurality of nodes (A, B, C, D) in a network comprising a closed loop configuration of one or more pairs of unidirectional transmission rings (1, 2) arranged to transmit data in opposite directions around the rings. Each node includes a respective message processor (5, 6) for each of the transmission rings (1, 2) and a host processor (60) linked to the message processors (5, 6). The traffic of data in each ring is dynamically monitored to obtain traffic information which is utilized by the message processors in accordance with a traffic control process to select one of the rings to transmit data from an originating node to a destination node. In the event of a fault in one of the rings, the other ring is utilized to transmit data at a reduced performance level while repairs are made to the faulty ring.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

Transmitting Data Between Multiple Computer Processors

The present invention relates to a method and apparatus for transmitting data between multiple computer processors.

In modern data communications involving multiple computer processors there are two traditional problems. Firstly, there is the unacceptable loss of data transmission speed when two or more processors attempt to communicate with each other. Secondly, multiple processor systems often show a complete or substantial failure after the occurrence of a transmission line problem.

Prior art techniques for transmitting data between processors or nodes have proposed recovery mechanisms in the event of a fault occurring in a ring connecting the processors, wherein data messages to be transmitted are looped back at a particular node and directed to an unaffected ring by a physical connection to an unaffected link at the node. None of the prior art techniques suggests a way of monitoring the data traffic on each of the rings linking the processors to select an optimum route for data traffic to travel to its destination processor or node.

It is therefore desirable to provide for an increase in available capacity enabling data transmission at an acceptable rate between processors in a ring network, and to provide for the transmission of data messages along the most expeditious route from an originating processor to a destination processor.

According to one aspect of the invention there is provided a method of transmitting data between a plurality of nodes containing computer processors, said method including the steps of:

connecting the nodes by a plurality of unidirectional transmission rings such that each ring is in a closed loop configuration, said transmission rings being arranged to transmit data in alternately opposed directions around the rings between the processors;

dynamically monitoring the traffic of data in each ring to obtain traffic information in each ring; and

utilising said traffic information to select one of the rings to transmit data in accordance with certain criteria.

The rings may be arranged in a layered configuration preferably comprising one or more pairs of unidirectional rings with each pair of rings being arranged to

transmit data in opposite directions.

Preferably, each node comprises a plurality of message processors, one for each transmission ring.

According to another aspect of the invention there is provided a communications system for transmitting data between a plurality of nodes in a network, comprising:

a closed loop configuration of two or more unidirectional transmission rings connecting the nodes, the transmission rings being arranged to transmit data between the nodes in alternately opposed directions around the rings;

each node including a respective message processor for each of the transmission rings;

wherein the message processors are programmed to select one of the rings to be used for transmitting a message from a node to another node in accordance with certain criteria.

Each node preferably includes a host processor which is linked to the message processors of the node.

When a data message is required to be transmitted from an originating node to a destination node, the host processor is preferably arranged to send the data message to each message processor associated with the originating node, and the message processors of the originating node select a ring on which the data is to be transmitted by utilizing the monitored information.

The traffic of data in each ring may be monitored to obtain information on any one or more of the following:

the available ring capacity;

the data flow rate or traffic loading on each ring; and

fault identification.

The message processors may perform their selection on the basis of information obtained from a look-up table. The look-up table may contain information about the number of ring links along which a data message has to travel along each ring between the nodes to reach its destination so that the shortest route for the data message can be determined. The look-up table may also contain

information about the data flow rate or traffic loading on each ring. Thus when one ring contains a lot of traffic and is congested, another ring may be selected. The look-up table is preferably dynamically updated for each new data message to be sent. For this purpose, counting means may be provided for counting the number of messages queued for transmission at a node or nodes of the system.

In accordance with another advantageous feature, a method in accordance with the invention may include the steps of determining whether data to be transmitted is priority data containing priority information and selecting one of the rings to transmit the priority data so as to provide the most expeditious route for the data to reach a destination node.

Packets of data containing priority information may contain a flag in a priority field to enable a message processor to determine that the data packet contains priority information. Packets of data having priority and queued for transmission may be transmitted ahead of packets queued for transmission that do not have priority.

In accordance with a further advantageous feature of the invention, one ring may be selected to transmit data of a particular kind and all other data is arranged to be transmitted on the other ring of a ring pair or, where there are more than two rings, on the other rings of the system. This is particularly useful when there is a large amount of data for a particular task to be transmitted from one node to another.

The method and system of the present invention may include means for performing maintenance functions, such as fault detection means for detecting when faults occurs in the transmission rings. In accordance with a preferred feature of the invention, when a fault is detected in one of the transmission rings, the system is arranged to transmit data messages only on the ring or rings not affected by the fault. This is in contrast to prior art techniques in which data messages are looped back at a node by a physical correction and directed onto an unaffected ring.

In accordance with another preferred feature the method and system of the invention utilize Scalable Coherent Interface (SCI) technology. The "Scalable Coherent Interface" is described in IEEE Standard P1596-1992 and other

publications including a paper entitled "The Scalable Coherent Interface and Related Standards Projects by David B. Gustavson (February 1992 – IEEE Micro, pp 10–21). The nodes in the system of the present invention preferably include scalable coherent interfaces (SCIs) which provide bus services by transmitting packets of data on print–to–print unidirectional links between the nodes. By using SCI technology the number of nodes and number of transmission rings in the method and system may be conveniently increased at any time by the addition of further SCIs.

In order that the invention may be more readily understood a particular embodiment will now be described, by way of example only, with reference to the accompanying drawings wherein:

Figure 1 is a schematic circuit block diagram of a communications system in accordance with the invention;

Figure 2 is a flow chart of a traffic control process used in the invention;

Figure 3 is a particular example of the diagram of Figure 1;

Figure 4 is a schematic block diagram showing the maintenance functions associated with a node;

Figure 5 is a frame structure for messages transmitted between message and most processors;

Figure 6 shows the maintenance (MA) information flow between a host processor and message processors at a node;

Figure 7 shows the transfer of maintenance information in the event of a fault occurring in one of the transmission rings of the system;

Figure 8 shows a fault recovery mechanism flow chart when a fault occurs;

Figure 9 is a block diagram of the main components of a message processor; and

Figure 10 is a block diagram showing the architecture of a NodeChip interconnection of the transmission rings and the message

processors of the system.

Referring to Figure 1, there is shown a topology of a Scalable Two-Way Ring (S2R) Structure comprising a loop 10 and four nodes, A to D, connected therein. The loop 10 comprises a pair of transmission rings 1 and 2 with each of the nodes A to D connected in the path of each ring 1 and 2. The structure has a scalable architecture which provides for multiple ring-layers so as to cope with various services, capacity and fault tolerances. The particular topology shown in Figure 1 is an example of a two-layer physical configuration to provide for services and single fault recovery over the same physical layer. In other words, the loop 10 may be considered as two identical ring layers, an inner ring 1 and an outer ring 2.

The loop 10 provides a bus service with packets that it transmits on point-to- point unidirectional links 21 and 22 between the nodes A to D. Each node, A to D, comprises two identical Interface Message Processors (IMPs), labelled 5 and 6, each being connected to a respective ring 11 or 12 of the loop 10. Each node may have more than two IMPs depending on the number of rings required. To deliver a message to its destination node, a host processor at an originating node is required to send the same message to all identical IMPs associated with each ring at the same originating node. The IMPs then decide which ring is to be used to send the message. The decision will be based on the information provided from a Dynamic Look-Up Table in accordance with a Traffic Control process, to be discussed with reference to Figure 2. The host will sequentially retrieve the message from each of the IMPs, e.g. 5 and 6, on the same node.

As indicated in Figure 1, the transmission paths of inner ring 1 and outer ring 2 are arranged in opposite directions. In normal operation, the transmission path of data in the inner ring 1 is in a clockwise direction from node A to node D. The transmission path of data in the outer ring 2 is in a counter-clockwise direction from node D to node A. With this two-way arrangement, packets can easily be routed between two adjacent nodes without going through the whole ring 1 or 2, to avoid traffic congestion. For example, when data is to be transmitted from node A to node B it may be transmitted along ring 1 and when data is to be transmitted

from node A to node D it may be transmitted along ring 2. However, data may be transmitted from node A to node D along ring 1 when there is less traffic in that ring than in ring 2. It is to be noted that any number of rings may be used, with the rings arranged in a layered structure, and that any number of nodes may be connected within each ring.

In order to handle the non-stop transmitting data between multiple computer processors over the network, a protocol, called the S2R protocol, has been developed on top of SCI protocols in the IMP. The S2R protocol will perform the functions of traffic control and data integrity control.

Traffic Control

In order to provide an efficient routing over the network, the following concepts will be employed:

- dynamic table of the traffic control process
- traffic balancing
- priority routing
- force ring.

Whenever a network is set up or a new node is introduced to the network, a dynamic table containing the following will be initiated:

- Start Node (S_n), being the originating node from which a message is to be transmitted;
- End Node (E_n), being the destination or termination node for the message;
- Ring Identity (R_{id}), corresponding to the rings on which the message can be transmitted;
- Node Cost (N_c), being the number of ring links a message has to pass through to reach the destination node;
- Traffic Loading (T_{ld}), being the number of messages queued for transmission;
- Combined Cost (C_c), being the sum of N_c and T_{ld};
- Next Ring Used (NR_u), being the next ring chosen to transmit a message, on the basis of a decision of the IMPs using the Traffic Control Process;
- Ring Total (R_t), being the total number of rings in the network;

- Maximum Traffic Load (TL_m), being a predetermined amount of traffic that the network can handle.

This table will be dynamically updated to reflect the amount of traffic in the network.

Therefore, by implementing the dynamic table of traffic control process, when a packet is received by a local traffic controller in an IMP, it will be able to select the most efficient routing. The steps that the algorithm uses can be demonstrated using the flow chart of Figure 2.

When a message arrives at step 200, the source address and destination address are set to S_n and E_n respectively, at step 202, to initiate the search of the dynamic table at step 204. A comparison is then performed at 206 to see if the combined cost, i.e. $C_c = N_c + T_{ld}$, has exceeded the limit of $TL_m + N_c$. If it has exceeded the limit the message is rejected at 208. For example, if the traffic loading T_{ld} exceeds the maximum traffic load TL_m, when a new message is to be transmitted, that message is rejected. If it does not exceed the limit a decision is made at 210 as to whether all entries returned from the dynamic table have the same combined cost, C_c. If they do have the same C_c, the traffic balancing concept is applied wherein a comparison is made to see if the Ring Identity, R_{id}, equals the Next Ring Used, NR_u, at 212. If it is the same, then that ring is used at 214, and for all entries returned, the traffic loading of that same ring is updated by incrementing the value of T_{ld} by 1, at 216. If R_{id} does not equal NR_u then the next ring used is updated by incrementing the value of NR_{ld} by 1 at 218. This will also occur for those returned entries that had $R_{id} = NR_u$. If the next ring used exceeds the total ring R_t at 220, the next ring used will be set to 1 at 222 and the process ends at 224. If the next ring used does not exceed R_t the process is stopped at 224.

If the combined costs, C_c, returned from all the entries are different at 210, then the route that has the minimum cost is chosen at 226. Once a ring is chosen for this route, the traffic loading of that ring is then incremented by 1 at step 228 and the process ends at 224.

An example of a dynamic look-up table that has real time updating during data transmission is shown in the following Tables 1(a) – 1(g) with reference to a

four node configuration shown in Figure 3. In Figure 3, the four nodes are labelled 61, 62, 66 and 69 and the outer and inner rings are labelled 11 and 12 respectively.

Table 1(a)

Initial Setup:

Entry	Start Node (S_n)	End Node (E_n)	Ring Identity (R_{id})	Node Cost (N_c)	Traffic Loading (T_{ld})	Next Ring (NR_u)	Combined Cost (C_c)
1	61	66	11	1	0	11	1
2	61	66	12	3	0	11	3
3	61	62	11	2	0	11	2
4	61	62	12	2	0	11	2
5	61	69	11	3	0	11	3
6	61	69	12	1	0	11	1

A transmission from 61 to 66, Ring #11 is chosen. The updated entries are:

Table 1(b)

1	61	66	11	1	1	11	2
2	61	66	12	3	0	11	3
3	61	62	11	2	1	11	3
5	61	69	11	3	1	11	4

A transmission from 61 to 66, Ring #11 is chosen. The updated entries are:

Table 1(c)

1	61	66	11	1	2	11	3
2	61	66	12	3	0	11	3
3	61	62	11	2	2	11	4
5	61	69	11	3	2	11	5

A transmission from 61 to 66, Ring #11 is chosen. The updated entries are:

Table 1(d)

Entry	S_n	E_n	R_{id}	N_c	T_{ld}	NR_u	C_c
1	61	66	11	1	3	12	4
2	61	66	12	3	0	12	3
3	61	62	11	2	3	11	5
5	61	69	11	3	3	11	6

A transmission from 61 to 66, Ring #12 is chosen. The updated entries are:

Table 1(e)

1	61	66	11	1	3	12	4
2	61	66	12	3	1	12	4
3	61	62	11	2	3	11	5
4	61	62	12	2	1	11	3
6	61	69	12	1	1	11	2

A transmission from 61 to 62, Ring #11 is chosen. The updated entries are:

Table 1(f)

1	61	66	11	1	1	12	2
3	61	62	11	2	1	12	3
4	61	62	12	2	0	12	2
5	61	69	11	3	1	12	4

A transmission from 61 to 62, Ring #12 is chosen. The updated entries are:

Table 1(g)

2	61	66	12	3	1	12	4
3	61	62	11	2	1	12	3
4	61	62	12	2	1	12	3
6	61	69	12	1	1	12	2

In Table 1(a) the initial set up shows two entries 1 and 2 for a message to be transmitted from start node 61 to end node 66 along rings 11 and 12 respectively, two entries 3 and 4 for a message to be transmitted from start node 61 to end node 62; and two entries 5 and 6 for a message to be transmitted from start node 61 to end node 6. Initially, the next ring to be used NR_u is the outer ring 11 for all entries 1 to 6. For entry 1 the node cost is 1, the traffic loading is initially zero so that the combined cost $C_c = N_c + T_{ld}$ is 1. For entry 2, the ring identity is ring 12, the node cost is 3 (as it passes along 3 links to get to node 66), the traffic loading is initially zero, so that $C_c = 3$.

When a message is required to be transmitted from node 61 to node 66, the next ring used NR_u, ring 11 is chosen and the updated entries are shown in table 1(b).

The traffic loading T_{ld} for each entry 1, 3 and 5 for ring 11 is incremented and so the combined cost C_c for those entries will also be incremented. As the C_c for entry 2 on ring 12 is more than the C_c for entry 1 on ring 11 (3 compared to 2 in table 1(b)), then from step 210 in Figure 2, the next ring to be used for transmitting a message from node 61 to node 66 is that with the minimum C_c, i.e. ring 11. That is $NR_u = 11$ and the traffic loading is accordingly incremented by one in Table 1(c). As no message is sent on ring 12, the conditions for entry 2 will remain unchanged, i.e. the node cost is still 3 and T_{ld} is still zero.

When the combined costs are compared for entries 1 and 2 in table 1(b), it is seen that C_c for entry 2 is still greater than C_c for entry 1 (3 to 2). Therefore ring 11 still has the minimum C_c and is chosen for the next transmission from node

61 to node 66 in table 1(c). The traffic loading for all entries 1, 3 and 5 for ring 11 is accordingly incremented by 1 to the value of 2 which increases the combined cost to 3. Again, no traffic is transmitted on ring 12 for entry 2 so its C_c remains at 3.

We now have the situation where the combined costs for both rings 11 and 12 are equal for entries 1 and 2. When another message is required to be transmitted from node 61 to 66, from step 210 of Figure 2, the process proceeds to step 212 to see if R_{id} equals NR_u. In this case (refer to table 1(c)), it does and so the message is transmitted on the same ring, i.e. ring 11 and the T_{ld} for ring 11 is incremented by 1 to the value of 3 as seen in table 1(d) for entry 1. The next ring used NR_u for entries 1 and 2 is then updated to ring 12 in table 1(d).

A new message to be sent from start node 61 to end node 66 will now be transmitted on ring 12 and the values of T_{ld} and C_c for entries 2, 4 and 6 are incremented as shown by Table 1(e).

For entries 2, 4 and 6 the traffic loading is updated to 1. The traffic loading for entry 1 remains at 4 as it has now changed rings.

To analyze a transmission from node 61 to node 62, it is necessary to consider initial values for entires 3 and 4 in table 1(a). From table 1(a) ring 11 is used for entries 3 and 4 and the combined costs are equal. For entry 3, $R_{id} = NR_u$ so ring 11 is used and the updated entries for T_{ld} and NR_u are shown in table 1(f). For entry 4, $R_{id} \neq NR_u$ and therefore just NR_u is updated by 1 to 12. When another message is required to be sent from start node 61 to end node 62, as seen in table 1(f), the combined cost for entry 3 is 3 which is greater than the combined cost for entry 4, which is 2. Therefore ring 12 is selected and the updated entries are shown in table 1(g). The T_{ld} for entry 4 is updated to 1.

When a packet of data is deemed to be important and requires immediate transmission, the priority routing concept can be employed. This will ensure that a packet with a high priority can by-pass the normal rule of routing and get to the destination as soon as possible. For example, if a packet is queued for transmission

at node 61 and intended for node 66 along inner ring 11, if it has priority over other packets queued ahead of it, then that priority packet may be transmitted to node 66 along ring 12 provided that this is the most expeditious path.

The force ring scheme can be used to delegate a task to a particular ring. That is, a selected ring will only be used to transmit particular specified messages while the other ring will carry the remainder of the traffic. It is particularly useful when there is a large amount of data required to transfer from one node in the network to another node.

Data Integrity Control

To ensure the message is transmitted correctly and accurately within the network, Message Verification and Message Sequencing will be utilized during the transmission.

A Checksum, Address Validation and Message Length Check will be used for the Message Verification. When a host processor sends a message to the IMPs, there must be a Checksum attached to the message. Each IMP makes its own calculation and compares it to the Checksum received in the message. If there is a mismatch, the message is discarded and an error signal will be issued to the host.

To ensure the Message Sequencing, a User-defined Flow Control (UFC) concept will be used with the following rules:

When a message is sent from the host to IMP, there are two services to be provided by IMP, i.e. Acknowledgement of message (AKM) and Retransmission of message (RTM). With these two services, three situations could happen:

(1) when the host does not require AKM, the IMP will continue to send the next message after sending the existing message.

(2) when the host requires AKM and RTM, the IMP will keep a copy of the message and retransmit the message when the Acknowledgement (ACK) is not received from the receiving node within a time-period (T_{ack}). This will be repeated until the maximum Resend Count (RC_m) is reached, in which case an error signal will be issued to the host.

(3) when the host requires AKM and no RTM, the IMP will send an error signal to the host if the timeout T_{ack} expires while awaiting for ACK response and

the IMP can continue to send next message.

Whenever a message is received in the IMP of the receiving node, a response ACK will be generated and returned to the sending node. If the message is faulty in some way, the message will be discarded and no response will be generated.

Both the value of the timeout T_{ack} and the RC_m are programmable from the host by setting the Control and Status Registers (CSR) in IMP.

Network Maintenance

The maintenance in the network is distributed and has a layered structure. Maintenance functions are carried out within each IMP of each node in relation to resources and parameters residing in the network's protocol entities. With reference to Figure 4, each IMP 5, 6 includes a S2R maintenance module 410 for performing S2R maintenance functions and an SCI maintenance module 420 for performing SCI maintenance functions. Between the S2R maintenance modules and the maintenance software in each host processor 60, there is established an S2R protocol which can implement the functions when necessary. A local processor bus protocol is established between the S2R and SCI maintenance modules 410 and 420. Between the transmission rings 11, 12 and the SCI maintenance modules, there is established an SCI protocol which again can implement the necessary functions when required. For each layer, the layer maintenance handles the specific maintenance information flows and provides the services to the upper layer.

In Figure 5 there is shown the frame structure 500 for a message transmitted between a host processor 60 and its associated IMPs 5, 6. The first field of bits 510 is reserved for the User-defined Flow Control (UFC), the coding and functionality of the bits being determined depending on the user application. The second field 520 is the destination address field, the bits indicating address data relevant to the destination node. The PT field 530 designates the Payload Type and is coded in 2 bits indicating the type of message including the data message, command message and the idle message. The Maintenance (MA) field 540 of 4 bits carries the information related to side identifier, fault and traffic status. The Priority (P) field 550 indicates whether or not a message has priority over other

messages to be transmitted. The Payload Field 560 contains the actual data to be transmitted or command data such as for the dynamic look-up table, or for the CSR during initialisation. The last field 570 is reserved for the Checksum for message verification.

Figure 6 shows information flow relating to the maintenance (MA) between the host and the message processors 5, 6. When a message is transmitted from the host 60 to any of message processors 5 and 6, packets 620 have the maintenance information bits (MA) 630 attached to them via multiplexer 610. The MA field is placed in the frame header resulting in the combined packet 640 being transmitted. On receiving a message, for example packet 650, the host 60 will extract or strip the MA bits 630 from each packet 650 and place the maintenance bits in the MA field of the next outgoing message.

SCI Maintenance Functions

All packets transmitted on the rings are covered by a Cyclic Redundancy Check (CRC). That is, any CRC errors are detected in each node and reported to S2R maintenance subsystem.

When sending a packet, the node will expect an acknowledgement to occur within a timeout period. If the sender does not receive the acknowledgement within this timeout period, it will increment the fault counter and cause the Status bit of Echo timeout to be asserted. Retransmission might then be done dependent on the application of the maintenance software.

When a ring is operational, synchronization packets will be sent on the down stream link within a given interval. If this interval becomes too long or the syncronization packets for some reason do not occur, it will cause a synchronization error to be flagged by the down stream node. Restart of the ring might then be performed dependent on the maintenance software. The restart sequence of the ring is handled by the SCI protocols.

S2R Self-Recovery Mechanism

The IMP defines a working mode and a protection mode. In normal operation, the IMP is configured in a working mode. If a fault X is detected, say on ring 12 of the network shown in Figure 7, the MA bits will be sent from IMP

5, connected in the faulty ring 12, via its host processor 60 to the other IMP 6 associated with each particular node so that transmission can resume on ring 11. In this way the IMP is reconfigured in a protection mode whereby all packets can be transmitted on the fault-free ring 11. This fault recovery mechanism is normally expected to be handled by S2R maintenance functions as shown in Figure 8. Under normal operation the maintenance functions monitor each IMP for faults at step 810, and if a fault 815 is detected at 815, the maintenance functions are invoked to reconfigure the IMP to the protection mode at 820. The IMP is re-initialized at 830 when repairs have been carried out to remove the fault.

The particular procedure will be as follows:

- If a signal to be transmitted fails, resend or re-transmit the signal. This will be handled by SCI maintenance functions.
- If the resending fails, initiate tests of the IMP and ring to locate faults, and then switch the traffic to the other ring. This will be handled by 2SR maintenance functions.
- If both fail, restart all IMPs. This shall be handled by maintenance software. There must be some routine test in place in each IMP, so that all IMPs can perform the restart if both rings fail.

Node installation or node replacement will not affect the normal traffic over the network. Each host will send a command message to IMPs to update the dynamic look-up table and CSR after the new node has been installed. The IMP of the new ring will send MA bits to the other side to take over the traffic, the old ring can then be disconnected and installed with the new IMP for the new node. After the new node is installed and attached to both rings, each IMP of the working ring (i.e. protection node) will gradually send MA bits to the other side of IMP to reconfigure both sides as working mode. The same procedure will also be applied to the node replacement except the update of the dynamic table.

Implementation

Each IMP, shown as 13 in Figure 9, comprises three main parts, a transmitter/receiver section 15, a S2R Protocol Controller (SPC) 16 and an SCI NodeChip 17.

The SPC 16 contains digital logic in a single Application Specific Integrated Circuit/Field Programmable Gate Array (ASIC)/(FPGA) which performs the protocol conversion functions between the NodeChip 17 and a host processor (not shown). The host processor communicates with IMP 13 through processor bus 14 and is specifically linked to the transmitter/receiver section 15 of the IMP 13.

Node-to-Node interconnection is implemented using the SCI NodeChip 17, which is a single-chip solution complaint with the physical and logical layers of the SCI standard as defined in the American National Standards Institute/Institute of Electrical and Electronics Engineers (ANSI/IEEE) Standard 1596-1992. The NodeChip is a Trade Mark of Dolphin Interconnect Solutions and its functions are explained in technical reference manual of the manufacturer.

The SCI NodeChip 17 is implemented in low-power, CMOS technology. It provides an input link 19 and output link 20 for unidirectional communication suitable for node-to-node ring topologies. A 64-bit bidirectional bus 18, called CBus, provides a communication path between the SCI NodeChip 17 and SPC 16. The link control unit 21 of NodeChip 17 comprises an input control 22 for receiving packets of data from other IMPs, an output control 23 for transmitting packets from its respective IMP to other IMPs on the same ring, and a bypass first in first out (FIFO) buffer 24 connected between each input control 22 and output control 23 of the NodeChips 17 associated with each IMP.

Figure 10 shows the architecture of an S2R loop having two ring layers 1 and 2 with three nodes A, B and C in which the output control 23 of a first NodeChip 17A is connected via a link 21 of a transmission ring to the input control 22 of the 17B associated with a neighbouring node B on the same ring layer 1 and so on until the ring is complete. The output control 23 of the NodeChip 17C of the last node C in the ring is linked to the input control of the first IMP NodeChip 17A. The bypass FIFO 24 is connected between the input control 22 and output control 23 of each NodeChip 17.

Buffer control 25 oversees the control of storing and queuing packets of data that have been received in RX buffer 26 and those packets stored and queued ready for transmission in the TX buffer 27. Each of the NodeChip 17 and SPC 16 has

a CBus Interface Unit 30 and 31 respectively for translating the packets and signals transmitted and received on CBus 18 into a format suitable for use respectively by the NodeChip 17 and SPC 16. The Control and Status Registers (CSR) 29 store data for carrying specified tasks within the IMP and the host.

The SPC 16 interfaces the SCI NodeChip 17 to the host processor and translates read and write transactions supported by the NodeChip 17 to transfer data between the host processor bus 14 and the remote S2R nodes. The protocol conversion functions between the NodeChip 17 and host processor are carried out under the control of S2R Protocol Control Unit 32. CBus control unit 33 oversees the control of data transmitted over and received from the CBus 18. FIFO buffers 34 and 35 stack the packets of data being transmitted to and received from the host and NodeChip 17 on a first-in first-out basis. The buffers are connected between CBus Interface Unit 31 and Bus Interface Unit 36 which receives and transmits the data packets to the TX/RX section 15.

A two-byte wide differential pseudo-ECL signal provides the link speed between the nodes of 125 Mbytes/s. To overcome the physical limitation of the node-to-node distance a Hewlett Packard G-Link HDMP-1000 parallel-to-serial chipset is used. The NodeChip can directly interface to this chipset to achieve 1 Gbit/s serial coaxial communication over distances of tens of metres.

As seen in Figure 10, the interconnection of each of the IMPs associated with a particular node is done through a processor bus 37, where each associated IMP is on a different ring layer. This enables each node to select the most appropriate ring to use to transmit a particular message.

The embodiment described hereinabove has disclosed a Scalable Two-Way Ring (S2R) architecture that uses the SCI technology to produce a highly reliable self-recovery ring system. A simple self-recovery procedure has been described based on the SCI protocols and leads to a rapid recovery from transmission line failure. The S2R protocol has the advantages of scalability, modularity, rapid self-recovery and real-time node installation and replacement. A dynamic traffic control algorithm has been described which enhances the utilisation of the dual-ring capacity. The user-defined flow control scheme handles the data sequencing while

force ring and priority routing schemes provide the user the flexibility of the ring system. Furthermore, the maintenance information flow scheme avoids the physical connections between the IMPs as well as providing a cost-effective transfer of maintenance information over the ring system. The described embodiment discloses a dual ring loop or system using a commercial SCI chipset. Clearly, because of its scalable architecture it can be designed in multiple loop layers to cope with various services, capacity and fault tolerance.

The dual ring architecture has the ability to recover rapidly from transmission line failure by having an alternative ring-layer and a simple recovery procedure. If one ring goes down the other will take over its work at reduced performance, but the system can still maintain a certain degree of traffic until the faulty part is fixed and brought back into operation. For military, banking, telecommunication and many other applications, the ability to continue operating in the face of hardware problems is of great importance.

Since modifications within the spirit and scope of the invention may be readily effected by persons skilled in the art, it is to be understood that the invention is not limited to the particular embodiment described, by way of example, hereinabove.

CLAIMS:

1. A method of transmitting data between a plurality of nodes containing computer processors, said method including the steps of:

connecting the nodes by a plurality of unidirectional transmission rings such that each ring is in a closed loop configuration, said transmission rings being arranged to transmit data between the nodes in alternately opposed directions around the rings;

dynamically monitoring the traffic of data in each ring to obtain traffic information in each ring; and

utilising said traffic information to select one of the rings to transmit data in accordance with certain criteria.

2. A method according to claim 1 wherein the rings are arranged in a layered structure and each node includes a plurality of message processors, one for each transmission ring.

3. A method according to claim 2 wherein each node includes a host processor linked to the message processors of the node.

4. A method according to claim 3 wherein when a host processor is required to transmit a data message from its originating node to a destination node, the data message is sent from the host processor to each message processor associated with that originating node and the message processors of the originating node select a ring to transmit the data on the basis of the monitored information.

5. A method according to claim 4 wherein said each message processor associated with the originating node performs its selection on the basis of information obtained from a look-up table in accordance with a traffic control process.

6. A method according to any one of claims 1 to 5 wherein said monitoring step includes monitoring each ring to obtain information on any one or more of the following: the available ring capacity; data flow rate on each ring; and monitoring of faults.

7. A method according to claim 6 wherein said selection is made in response to any one or more of the following: the available ring capacity; data flow rate on each ring; and fault identification.

8. A method according to any one of the preceding claims wherein said method utilizes Scalable Coherent Interface (SCI) technology.

9. A method according to any one of the preceding claims wherein the transmission of data messages between the nodes is controlled by a protocol.

10. A method according to claim 9 wherein the protocol controls the traffic of data in each of the transmission rings and controls the integrity of the data transmission between the computer processors of the nodes.

11. A method according to claim 10 wherein the protocol is implemented in each of the processors of each node and controls the selection of a ring on which to transmit data messages, said selection being made on the basis of information obtained from a look–up table in accordance with a traffic control process.

12. A method according to claim 5 or claim 11 wherein the look–up table is dynamically updated for each new data message to be sent.

13. A method according to any one of the preceding claims wherein the traffic loading on each ring is used to determine the ring that is selected to be used to transmit a data message.

14. A method according to any one of the preceding claims wherein the number of ring links along which a data message has to travel between nodes to reach its destination is used to determine the ring that is selected to be used to transmit the data message.

15. A method according to any one of the preceding claims wherein the processors are arranged to carry out maintenance functions.

16. A method according to claim 15 wherein, in the event of a fault occurring on one ring, the data messages are transmitted only on the ring or rings not affected by the fault.

17. A method according to claim 16 wherein, in the event of a fault occurring in one ring, maintenance bits associated with data packets being transmitted or queued for transmission on the faulty ring, are transferred to other processors at each node so that transmission of the affected packets can continue on other rings not affected by a fault.

18. A method according to any one of the preceding claims comprising the further steps of determining whether data to be transmitted is priority data containing priority information and selecting one of the rings to transmit said priority data so as to provide the most expeditious route for said priority data to reach the destination node.

19. A method according to any one of the preceding claims further comprising the steps of selecting one ring on which to transmit data of a particular kind and transmitting all other data on another ring or other rings.

20. A method of transmitting data between a plurality of nodes containing computer processors, said method including the steps of:

connecting the nodes by a plurality of unidirectional transmission rings, each

ring being in a closed loop configuration, said transmission rings being arranged to transmit data around the rings between the nodes in alternately opposed directions;

determining whether data to be transmitted contains priority information; and

selecting one of the rings to transmit said data so as to provide the most expeditious route for the data to reach a destination node.

21. A method according to claim 18 or claim 20 wherein said determining step is performed by reading packets of data to see if a priority field in the packets is flagged indicating that it has priority.

22. A method according to claim 21 wherein packets of data having priority and queued for transmission will be transmitted ahead of packets queued for transmission that do not have priority.

23. A method of transmitting data between a plurality of nodes containing computer processors, said method including the steps of:

connecting the nodes by a plurality of unidirectional transmission rings, each ring being in a closed configuration and said transmission rings each arranged to transmit data in alternately opposed directions around the rings between the nodes;

selecting one ring on which to transmit data of a particular kind; and

transmitting all other data on another ring or other rings.

24. A communications system for transmitting data between a plurality of nodes in a network, comprising:

a closed loop configuration of two or more unidirectional transmission rings connecting the nodes, the transmission rings being arranged to transmit data between the nodes in alternately opposed directions around the rings;

each node including a respective message processor for each of the transmission rings;

wherein the message processors are programmed to select one of the rings to be used for transmitting a message from a node to another node in accordance

with certain criteria.

25. A communications system according to claim 24 wherein each node contains a host processor which is linked to the message processors of the node.

26. A communications system according to claim 24 or claim 25 wherein the host processor at an originating node is arranged to send a data message to each of the message processors at the originating node, and the message processors then select which ring is to be used to send the message.

27. A communications system according to claim 26 wherein the message processors at an originating node are programmed to select the ring to be used on the basis of information obtained from a look-up table.

28. A communications system according to claim 27, wherein the look-up table is dynamically updated for each new data message to be sent.

29. A communications system according to any one of claims 24 to 28 including fault detection means for detecting when faults occur in the transmission rings.

30. A communications system according to claim 29 wherein when a fault is detected in one of the transmission rings, the system is arranged to transmit data messages only on the ring or rings not affected by the fault.

31. A communications system according to any one of the preceding claims wherein the transmission rings are arranged in a layered configuration of at least one pair of unidirectional rings arranged to transmit data in opposite directions around the rings.

32. A communications system according to any one of claims 24 to 31 wherein each message processor comprises a scalable coherent interface.

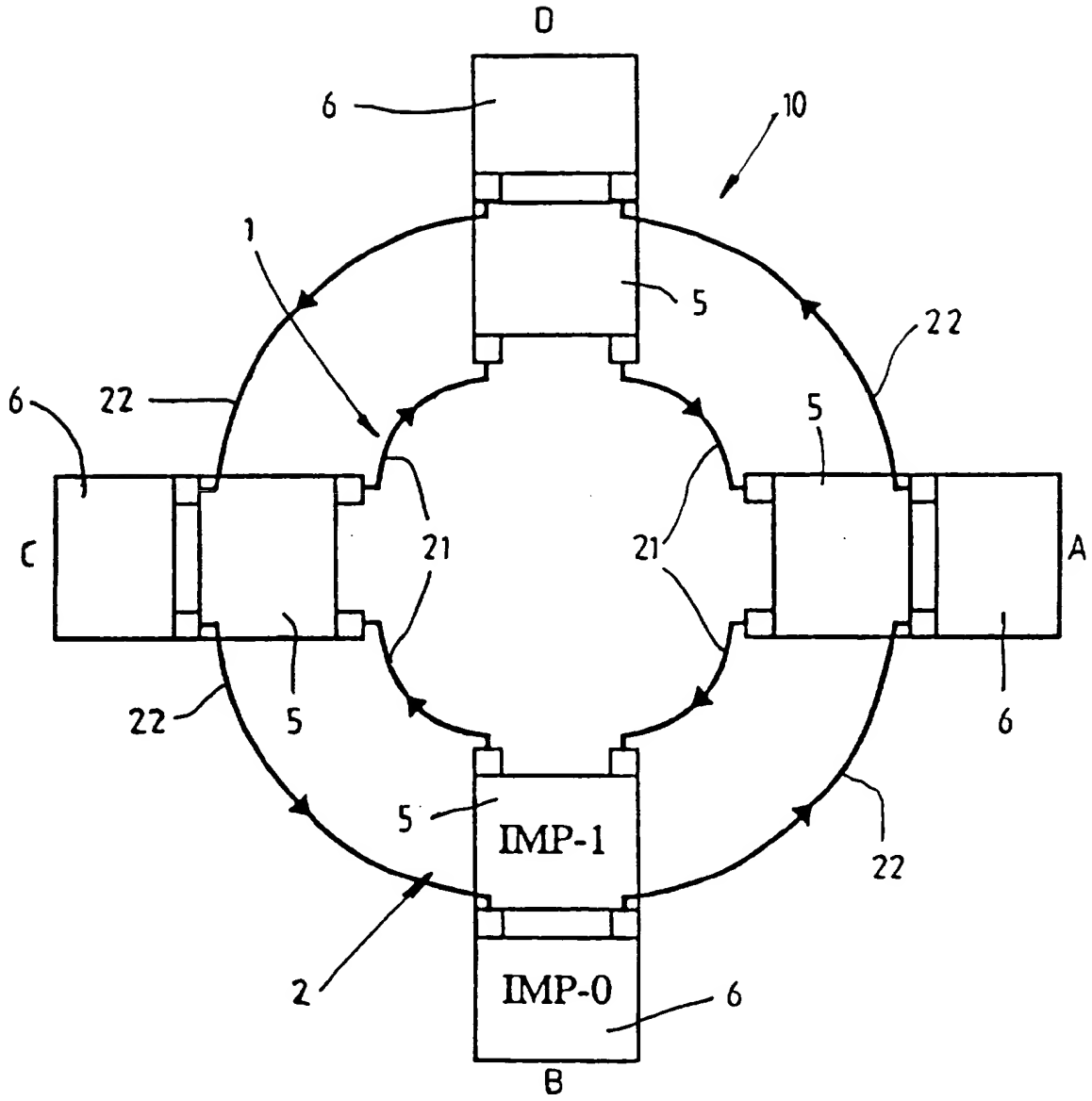


FIG. 1.

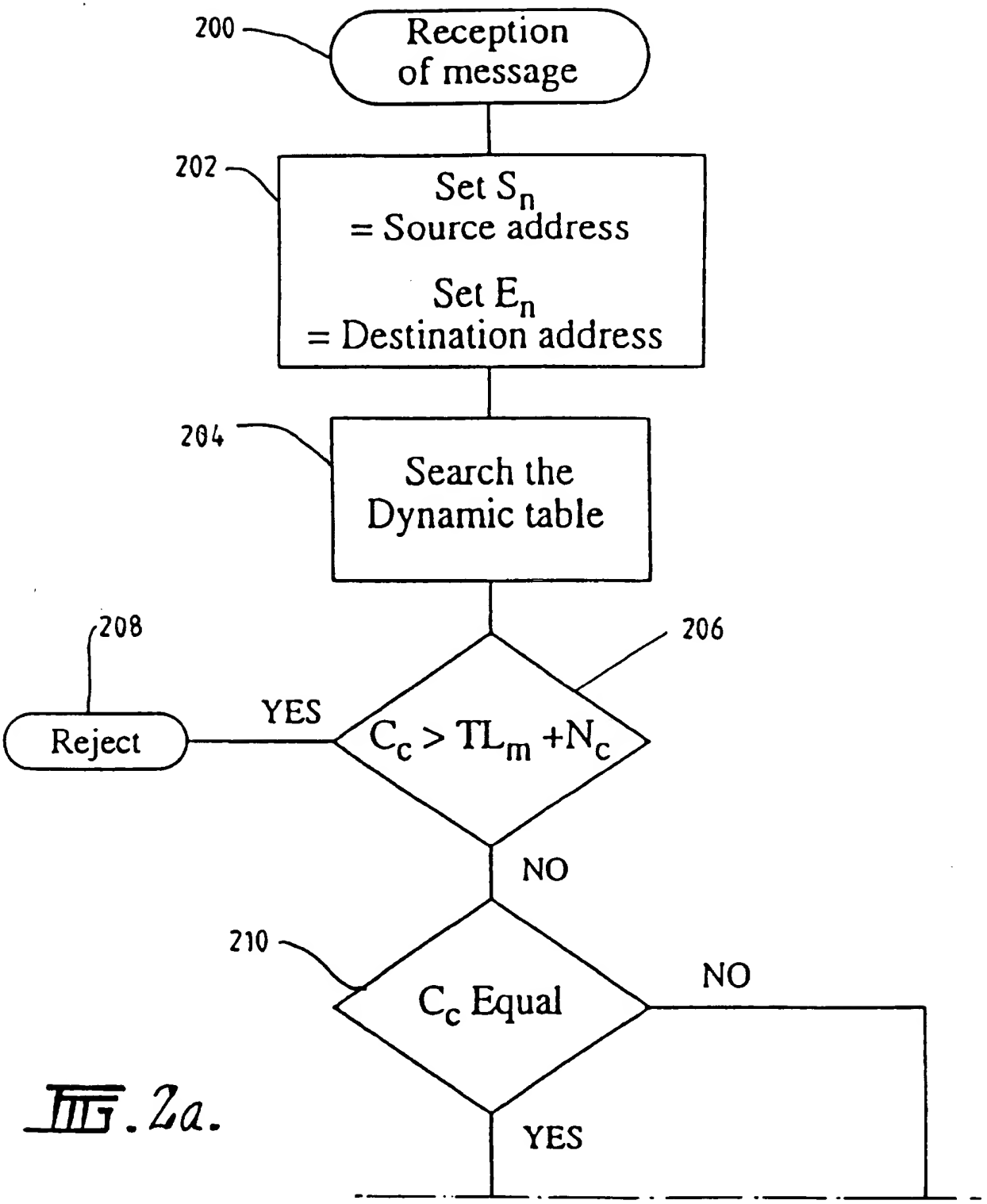

200
Reception of message
202
Set S_n = Source address
Set E_n = Destination address
204
Search the Dynamic table
206
208
Reject
YES
$C_c > TL_m + N_c$
NO
210
C_c Equal
NO
YES

FIG. 2a.

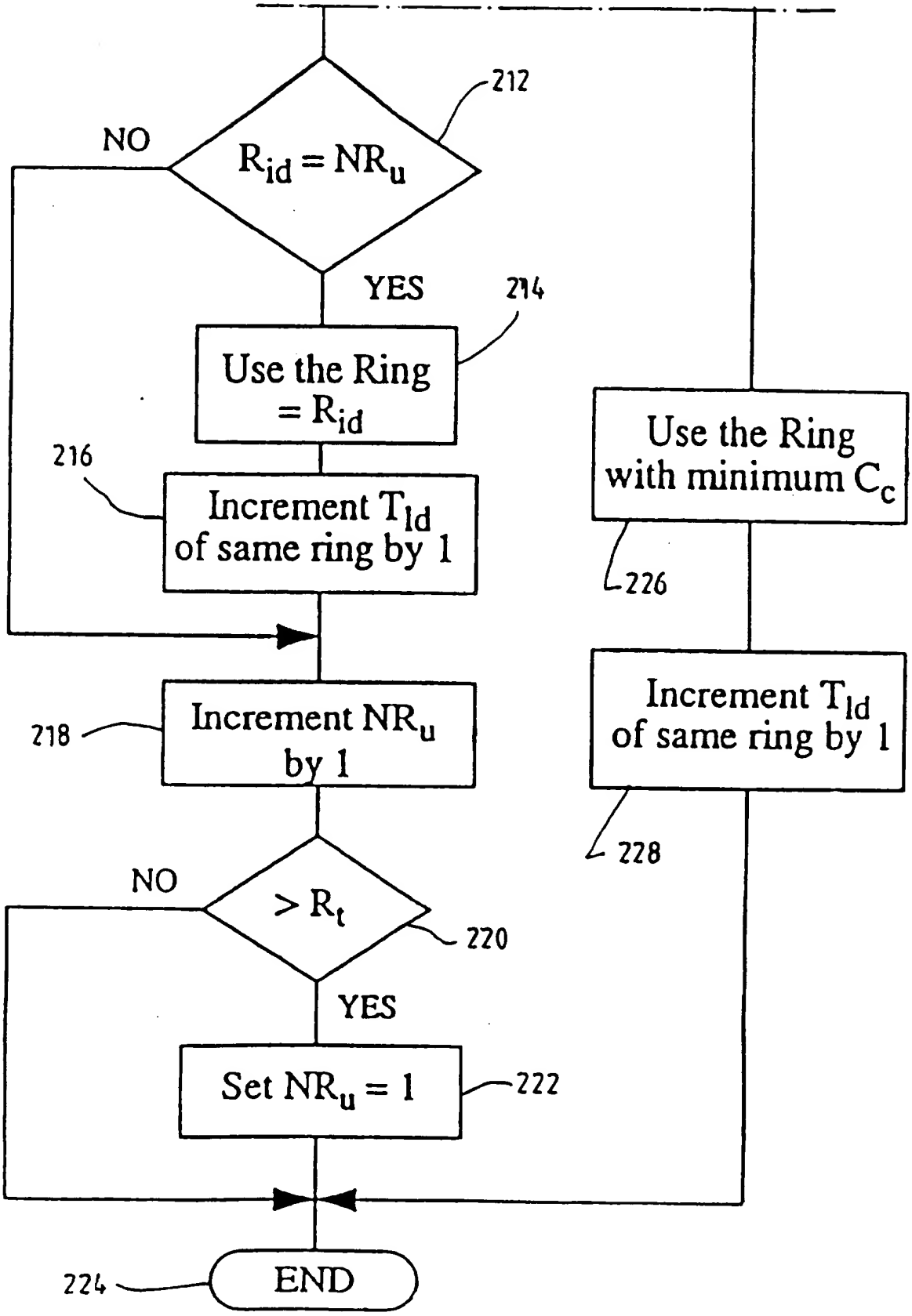

212
NO
$R_{id} = NR_u$
YES
214
Use the Ring
$= R_{id}$
216
Increment T_{Id}
of same ring by 1
Use the Ring
with minimum C_c
226
Increment T_{Id}
of same ring by 1
228
218
Increment NR_u
by 1
NO
$> R_t$
220
YES
Set $NR_u = 1$
222
224
END

FIG. 26.

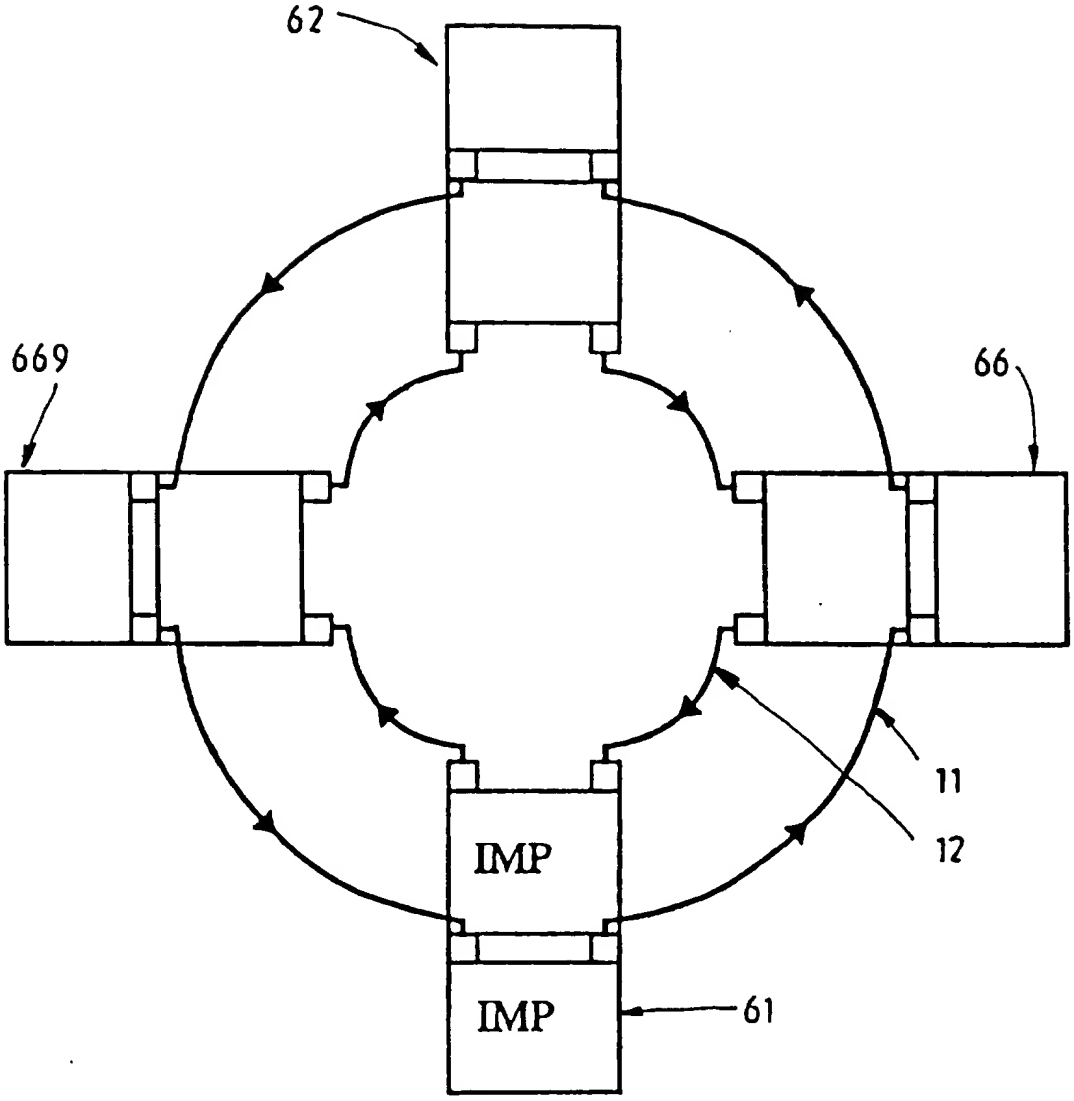


FIG. 3.

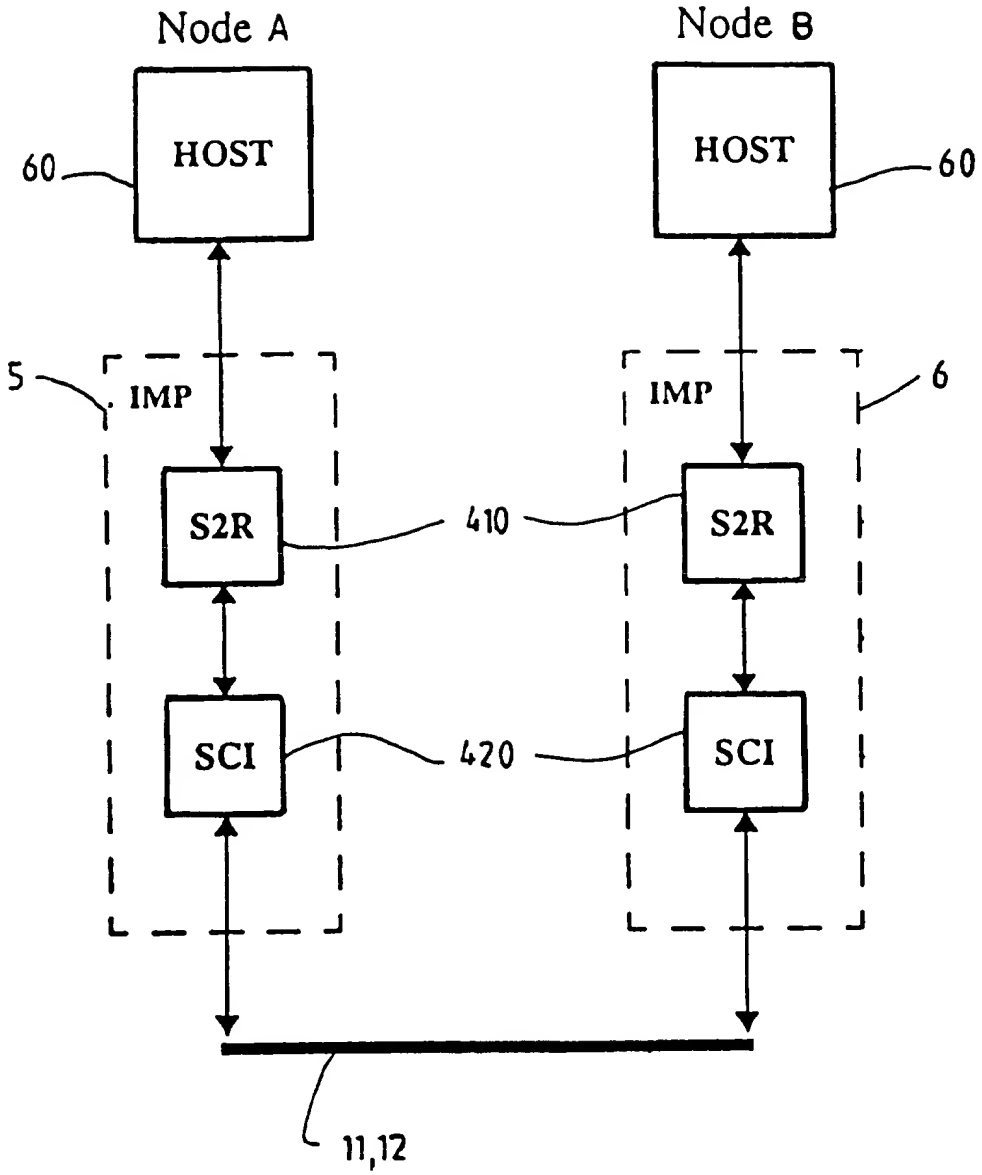

Node A
Node B
HOST
HOST
60
60
5
IMP
IMP
6
S2R
S2R
410
SCI
SCI
420
11,12

FIG. 4.

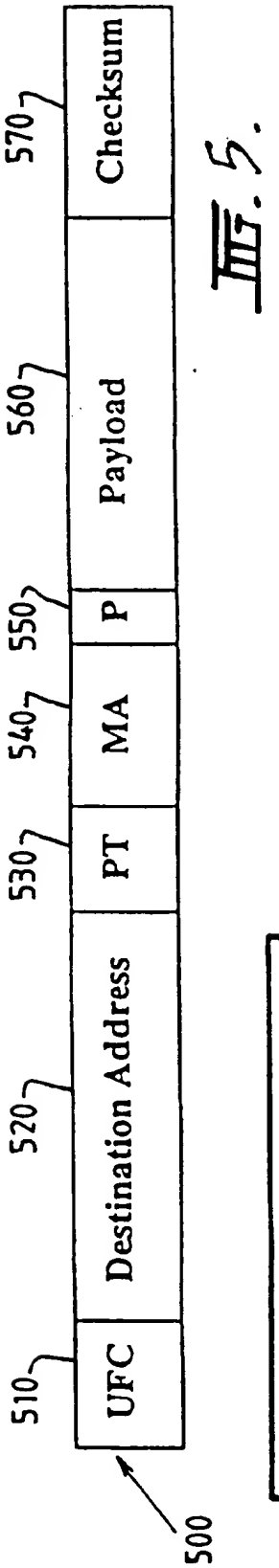

UFC
Destination Address
PT
MA
P
Payload
Checksum
510
520
530
540
550
560
570
500

Fig. 5.

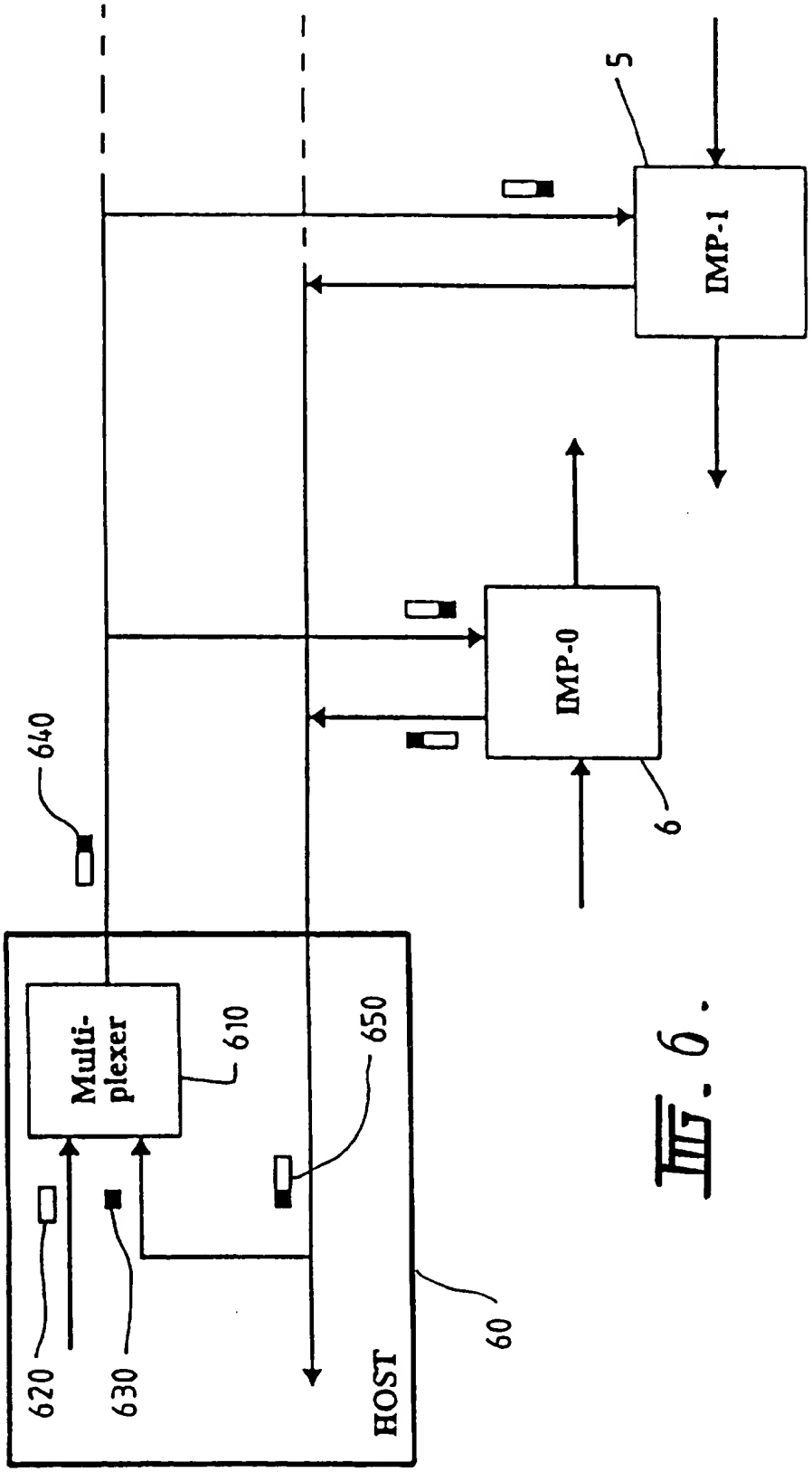

IMP-1
IMP-0
Multi-plexer
HOST
610
620
630
640
650
5
6
60

Fig. 6.

Fig. 7.

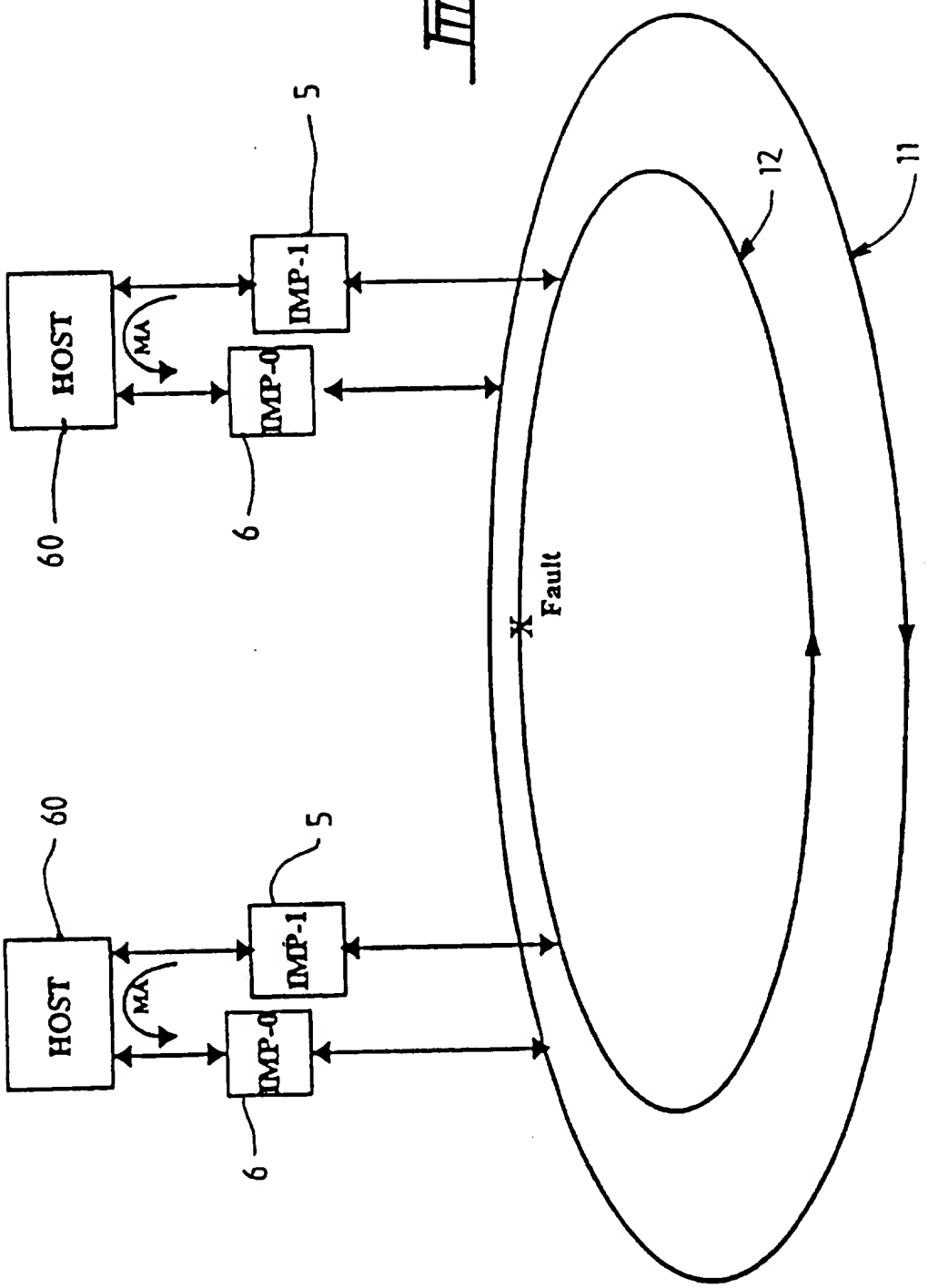
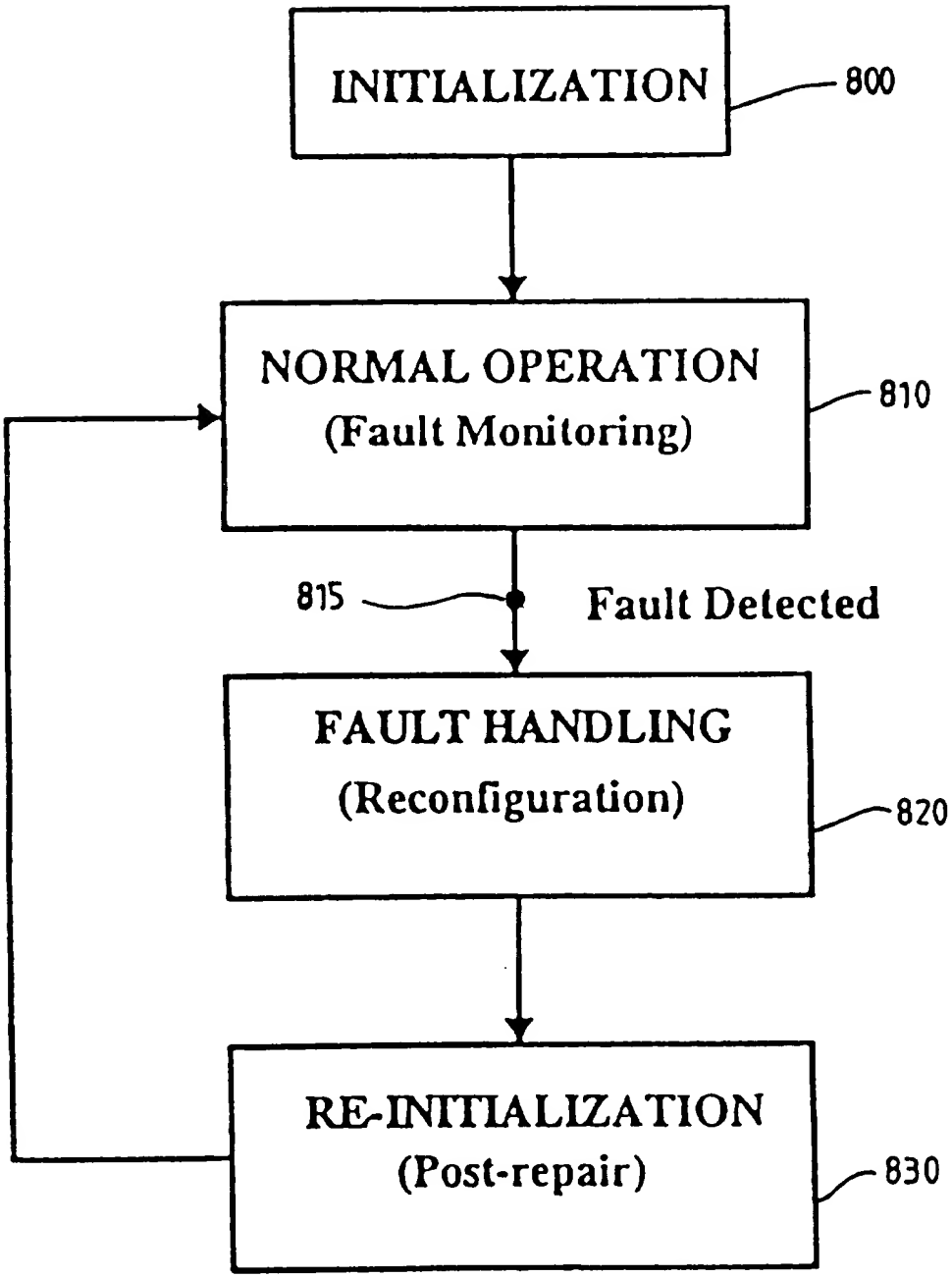


FIG. 8.


INITIALIZATION
800
NORMAL OPERATION
(Fault Monitoring)
810
815
Fault Detected
FAULT HANDLING
(Reconfiguration)
820
RE-INITIALIZATION
(Post-repair)
830

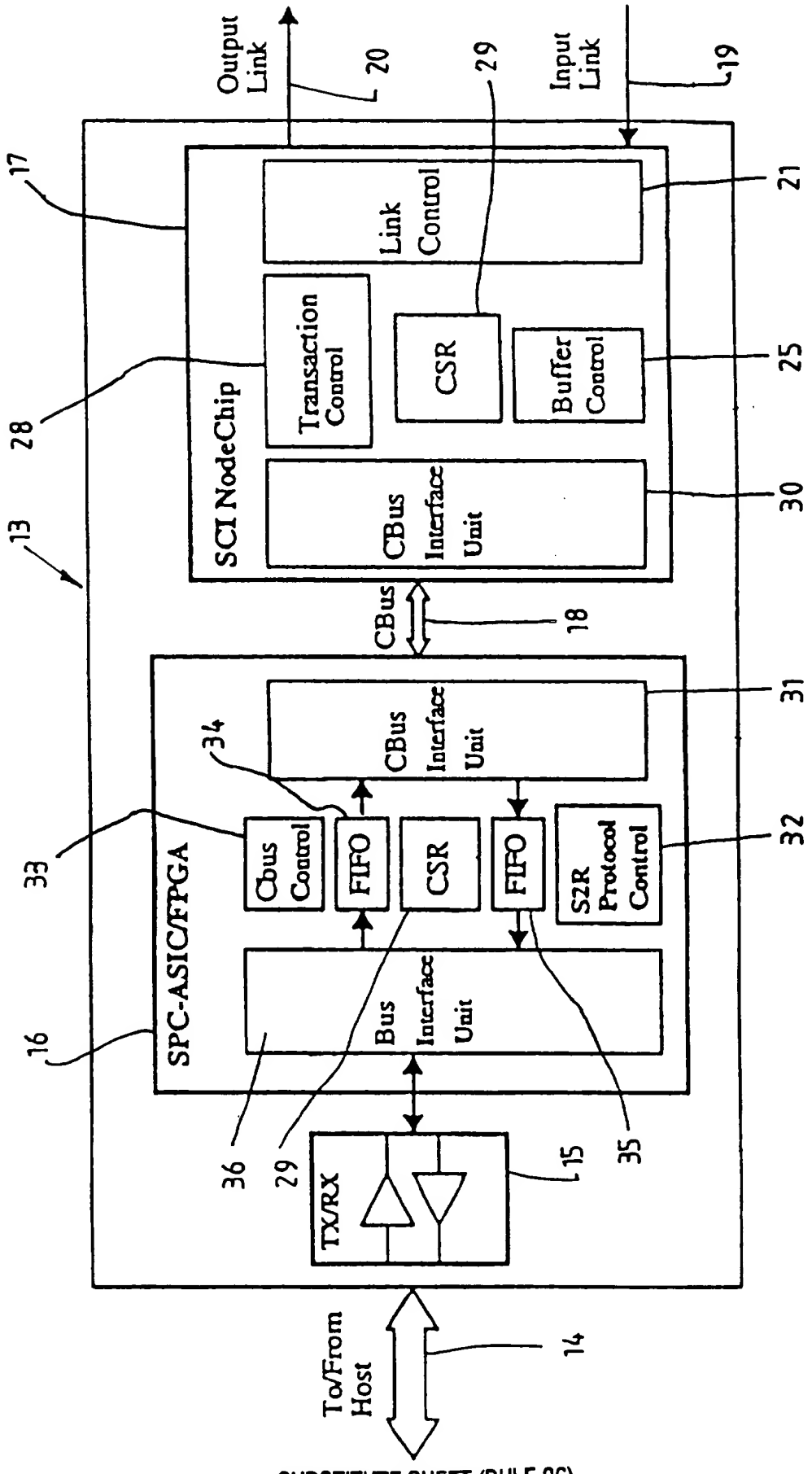

Output Link
20
29
Input Link
19
17
21
Link Control
Transaction Control
CSR
Buffer Control
25
28
SCI NodeChip
CBus Interface Unit
30
13
CBus
18
31
CBus Interface Unit
34
33
Cbus Control
FIFO
CSR
FIFO
S2R Protocol Control
32
SPC-ASIC/FPGA
16
Bus Interface Unit
36
29
TX/RX
15
35
To/From Host
14

FIG. 9.

Node.A.

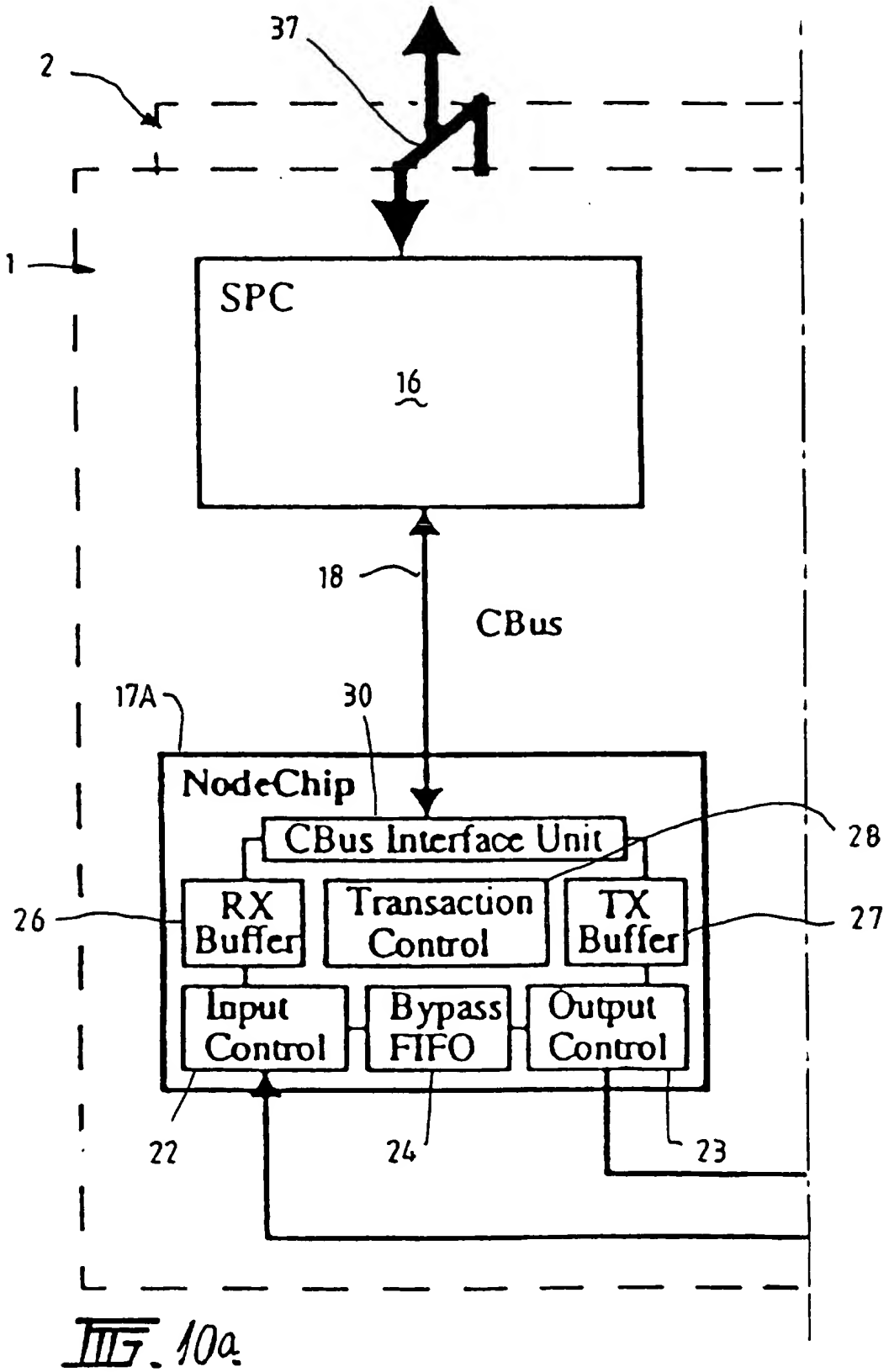


FIG. 10a.

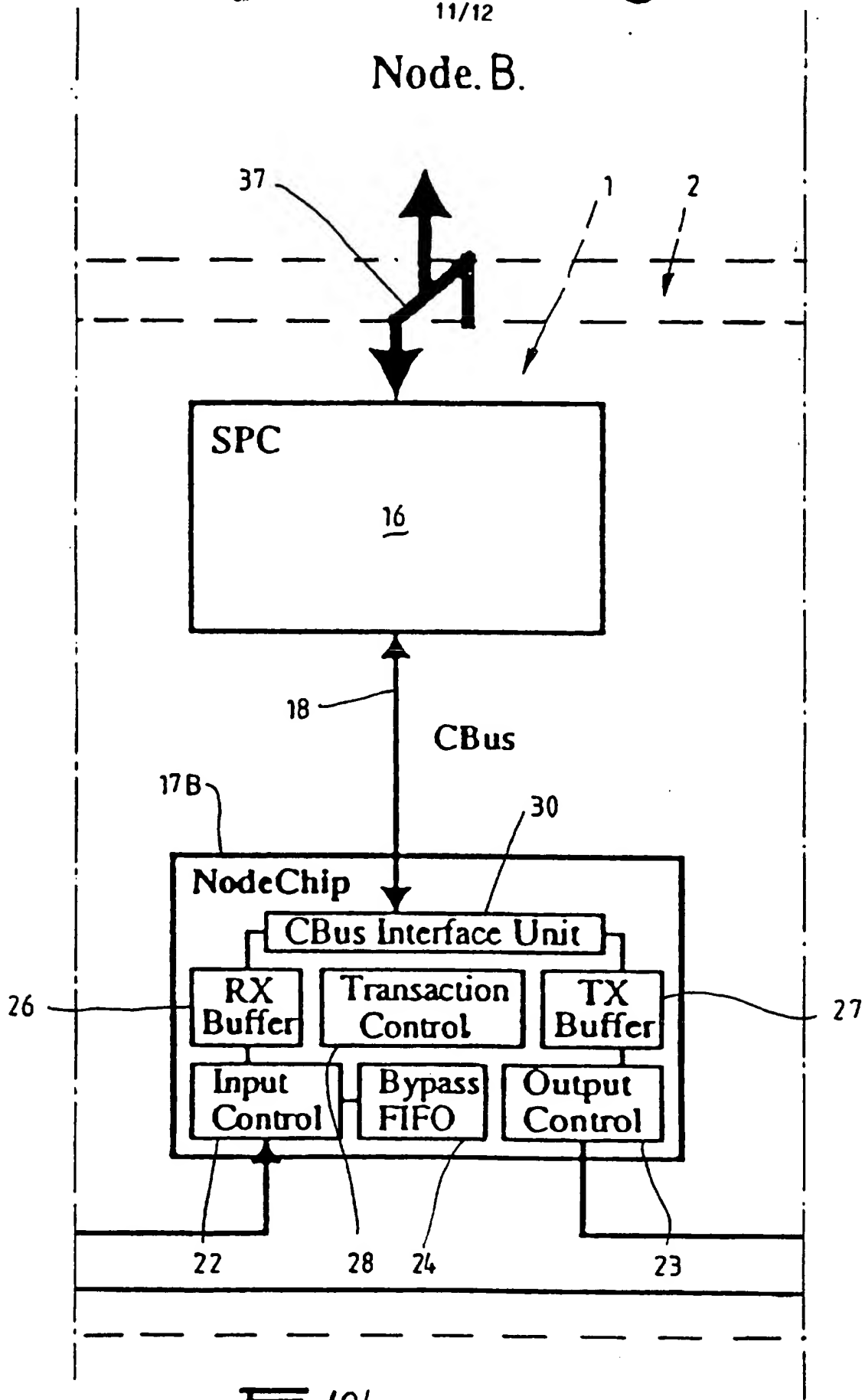

Node. B.
37
1
2
SPC
16
18
CBus
17B
30
NodeChip
CBus Interface Unit
RX Buffer
Transaction Control
TX Buffer
26
27
Input Control
Bypass FIFO
Output Control
22
28
24
23

FIG. 10b

Node.C.

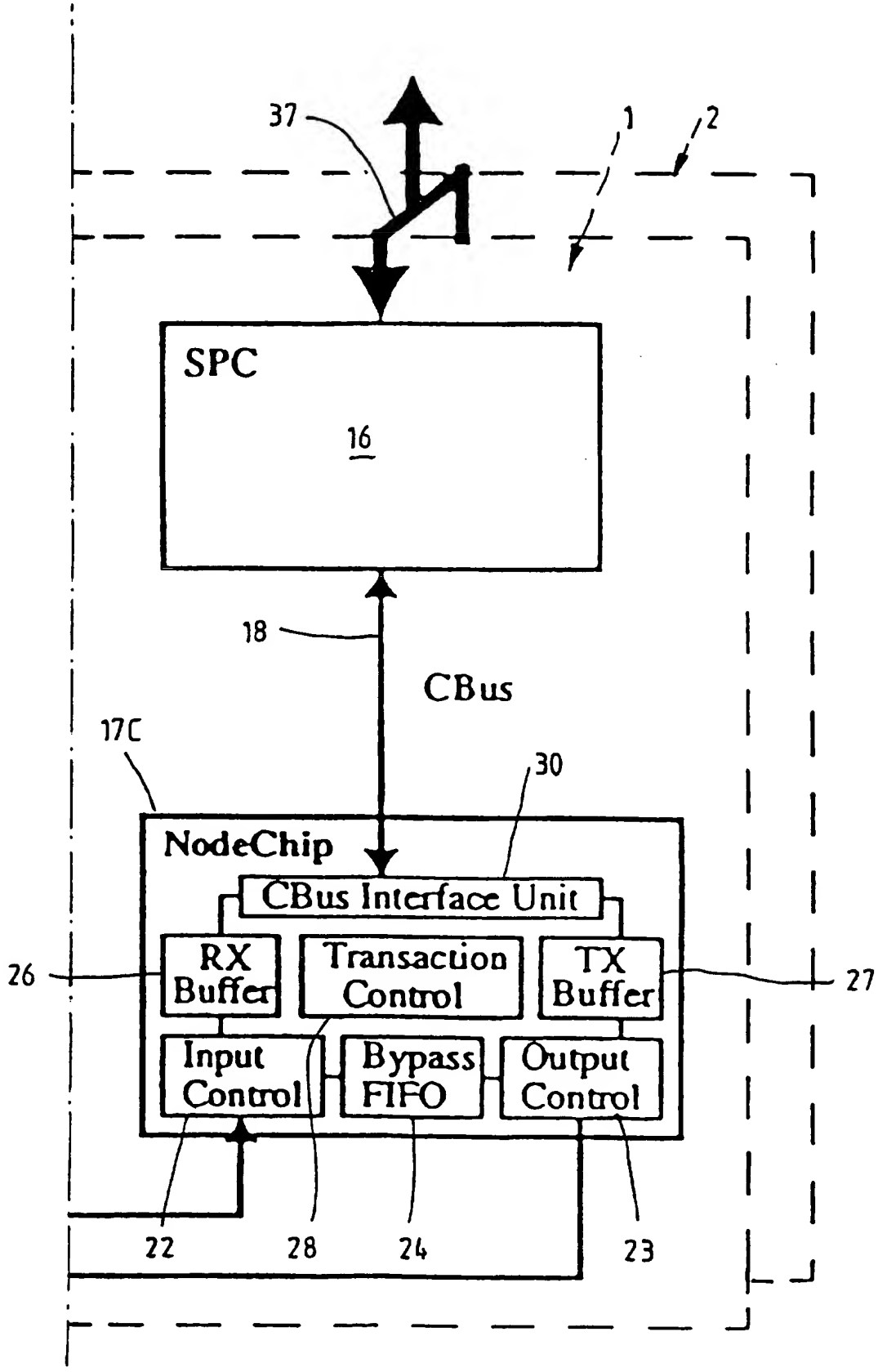


FIG. 10C

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/AU 96/00621

A. CLASSIFICATION OF SUBJECT MATTER

Int Cl[6]: H04L 12/42, H04L 12/437, G06F 13/40, G06F 15/173

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC : as above

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
AU : IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
WPAT: INSPEC (FDDI, DUAL BUS, SCI or SCALABLE)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	EP 590872 (AT & T) 7 January 1994 Abstract	1,2,11-13,31 3-5
X	US 5282199 (IBM) 25 January 1994 Whole document	1,6,7,9,10
X Y	WO 93/00756 (Bell Communications Res.) 7 January 1993 Whole document	1,2,3,31 8,32

[X] Further documents are listed in the continuation of Box C

[X] See patent family annex

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 October 1996	30 OCT 1996
Name and mailing address of the ISA/AU AUSTRALIAN INDUSTRIAL PROPERTY ORGANISATION PO BOX 200 WODEN ACT 2606 AUSTRALIA Facsimile No.: (06) 285 3929	Authorized officer **DALE SIVER** Telephone No.: (06) 283 2196

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 471633 (France Telecom) 19 February 1992 Whole document	1,9,10,15,16, 18-24,29,30
X	EP 468813 (NEC) 29 January 1992, Abstract, figures	1,6,7,14-17
X	"The QPSX MAN" (Newman et al) April 1988 - Vol. 26, No. 4 pp20-28, IEEE Communications Magazine Whole document	20-22
X	EP 158364 (Unisearch) 16 October 1985 Whole document	20-22
Y	US 5351040 (Matsuura et al) 27 September 1994 Abstract, figures	15-17
Y	"Connecting the AP1000 with a Mainframe for computation of the Experimental High Energy Physics" (Ichikawa et al) March 1993 - Fujitsu Sci. Tech. J. Vol. 29, 1, pp97-111 Abstract, sec. 2.3, figures 1-3	2,3
Y	"The Scalable Coherent Interface and Related Standards Projects" (Gustavson) February 1992 - IEEE MICRO, pp10-21	8,32

INTERNATIONAL SEARCH REPORT
Information on patent family members

International Application No.
PCT/AU 96/00621

This Annex lists the known "A" publication level patent family members relating to the patent documents cited in the above-mentioned international search report. The Australian Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent Document Cited in Search Report		Patent Family Member					
US	5351040	JP	4114535				
EP	590872	CA	2099972	JP	6205028	US	5406401
US	5282199	JP	6237261				
WO	9300756	CA	2112386	EP	591429	JP	6508967
		US	5179548				
EP	471633	FR	2665967	JP	6019820		
EP	468813	CA	2047949	DE	69114203	JP	4084535
		US	5150356				
EP	158364	AU	40998/85	US	4663748		

END OF ANNEX

(12) UK Patent Application (19) GB (11) 2 133 952 A

(21) Application No 8333570
(22) Date of filing 16 Dec 1983
(30) Priority data
(31) 2/59974
(32) 22 Dec 1982
(33) Belgium (BE)
(43) Application published
1 Aug 1984
(51) INT CL³
H04L 11/16
(52) Domestic classification
H4P PL
(56) Documents cited
GB 1510462
GB 1168476
(58) Field of search
H4K
H4P
G4H
(71) Applicant
International Standard Electric Corporation (USA—Delaware), 320 Park Avenue, New York 10022, State of New York, United States of America
(72) Inventors
Pierre Leon Debuysscher, Hugo Jan Philemon Peeters, Christiaan Georges Madeleine Maria Hennebel
(74) Agent and/or Address for Service
P. G. Ruffhead, ITT Patent Department UK, Maidstone Road, Foots Cray, Sidcup DA14 5HT

(54) **Multiple-ring communication system**

(57) A communication system with a plurality of nodes (Nl to Nn) intercoupled in two rings (RG1, RG2) operating in reverse directions on an equal basis and without central control. Each node is able to transmit on these rings, ring test messages the destination of which is the node itself, as well as node test messages which, upon receipt in the neighbouring nodes on a ring give rise to the transmission of reply messages on the other ring.

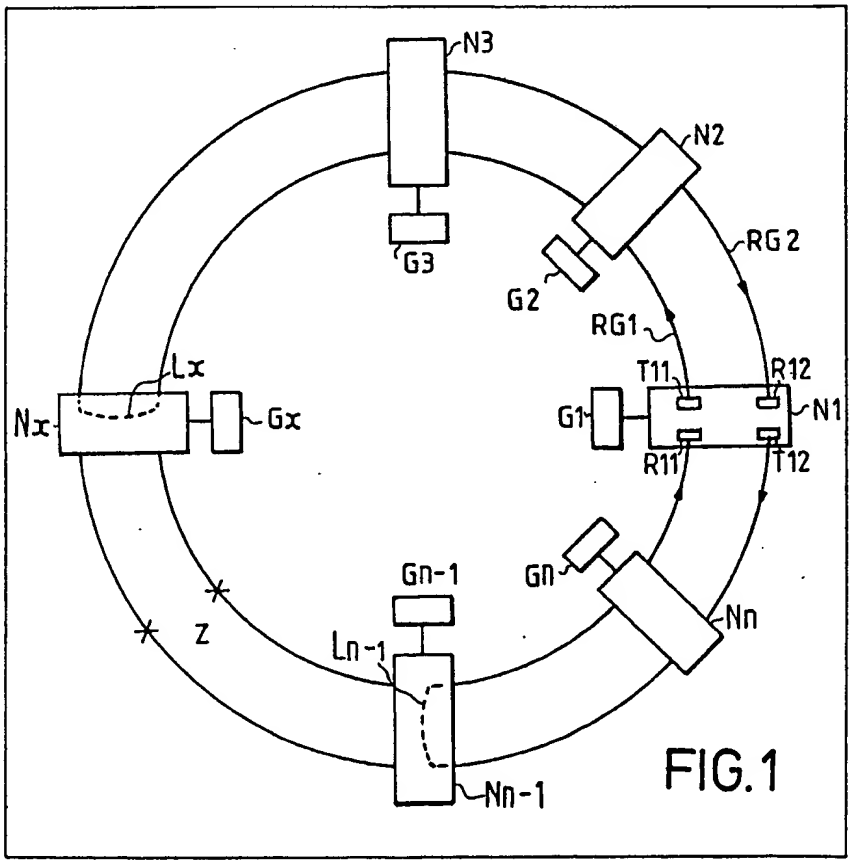


FIG.1

GB 2 133 952 A

The drawings originally filed were informal and the print here reproduced is taken from a later filed formal copy.

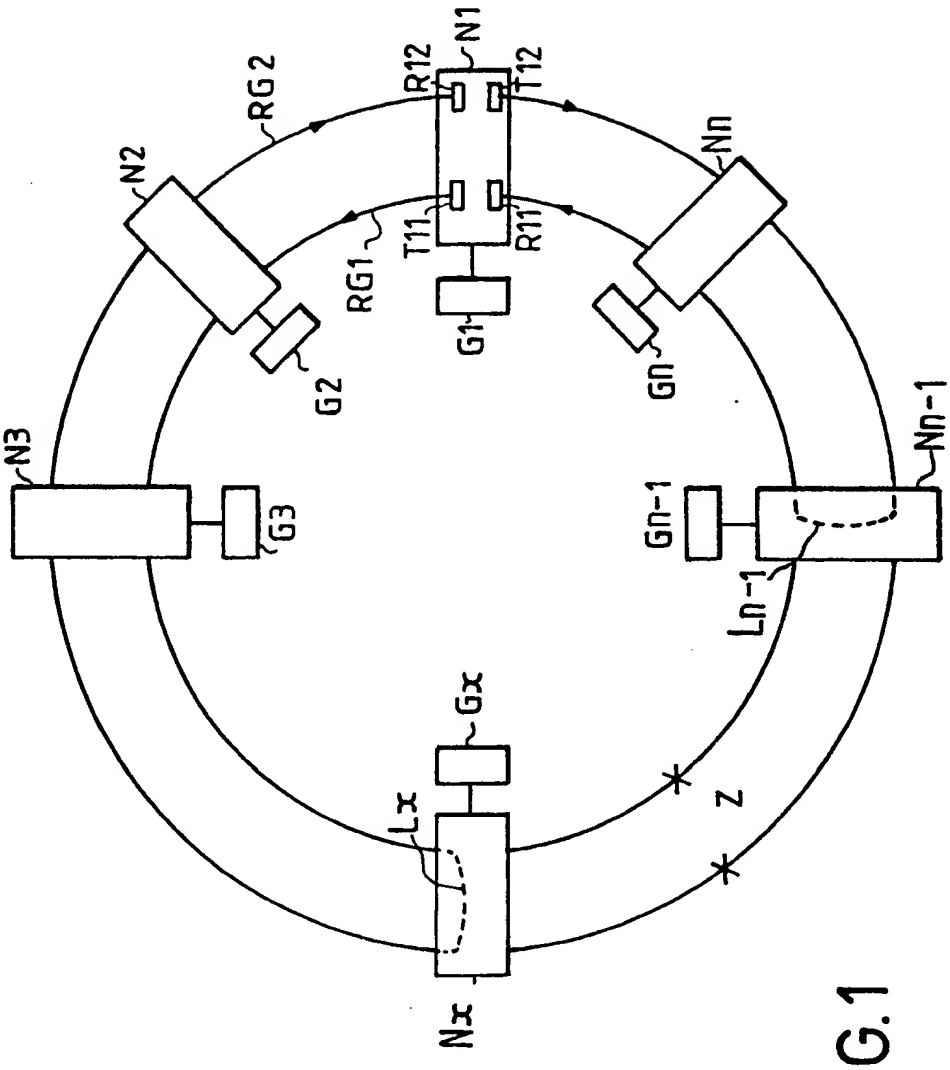

N1
R12
T12
RG2
N2
Nn
T11
R11
RG1
G1
G2
Gn
N3
Nn-1
G3
Gn-1
Ln-1
Gx
Lx
Z
Nx

FIG.1

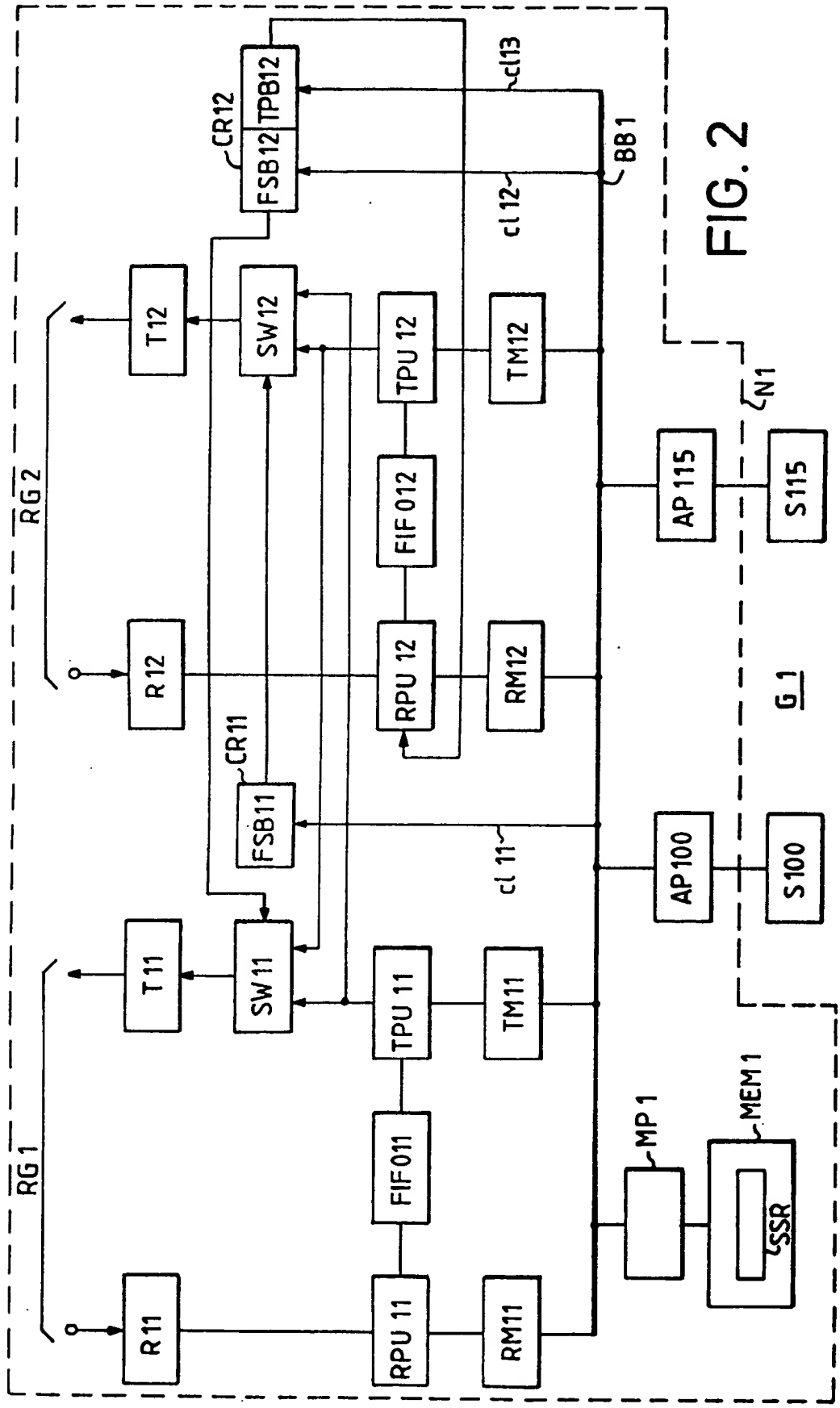

RG 1
RG 2
R11
R12
T11
T12
RPU 11
RPU 12
RM11
RM12
TPU 11
TPU 12
TM11
TM12
SW11
SW12
FIF011
FIF012
FSB11
FSB12
TPB12
CR11
CR12
cl 11
cl 12
cl13
BB1
AP100
AP115
S100
S115
N1
G 1
MP 1
MEM 1
SSR

FIG. 2

SPECIFICATION

Multiple-ring communication system

This invention relates to a multiple-ring communication system of the kind which includes a plurality of nodes intercoupled in at least two rings and operating on an equal basis without central control, each node including a plurality of message receiver/transmitter equipments associated with distinct ones of the rings respectively and able to receive messages on the rings and transmit messages thereon in opposite directions, and processing means for checking the receipt or absence of receipt of signals on the receiver ring portions connecting the receivers of the node to transmitters of its neighbouring nodes and for performing reconfiguration operations in response to the result provided by the checking operation.

Such a system is already known from the article "A distributed double-loop computer network (DDLCN)" by J.J. Wolf and M.T. Liu, published on pages6—19 to 6—34 of the Proceedings of the Seventh Texas Conference on Computing Systems, held in Houston, October 30—November 1, 1978.

In order to have a maximum numberof nodes always able to communicate with each other, even when there is some fault in the system, it is necessary to test the condition of the system almost continuously and, in the case where an error is detected, to take immediately the necessary reconfiguration measures to restore the faulty system to an operative status. Such a reconfiguration measure consists, for instance in the case of an interruption of both rings, in transforming the double ring system into a single ring system by establishing two interconnections between these rings.

In the above mentioned known system each node transmits to its neighbouring nodes timing signals as long as it is not transmitting messages, so that each node receives such timing signals in a substantially continuous way on the receiver ring portions by means of which its receivers are coupled to the transmitters of the neighbouring nodes. Each node tests the condition of its neighbouring nodes by checking the receipt or absence of receipt of these timing signals. In this way, very limited information is obtained at the testing node on the overall status of the system since this information only concerns the condition of the receiver ring portions, of the transmitters of the neighbouring nodes and of its own receivers. No information at all is obtained at this node on the condition, for instance of its own transmitters, of the receivers at the neighbouring nodes and of the transmission ring portions by means of which its transmitters are coupled to receivers at the neighbouring nodes. As a consequence, reconfiguration measures have to be taken by the processing means on the basis of very limited information on the system status and these measures can therefore be erroneous. For instance, it might be decided to establish two loops between the rings to form a single ring system although one of the rings is still fully operational.

An object of the present invention is to provide a multiple-ring communication system of the above kind, but each node of which is adapted to collect more information on the status of the system.

According to the invention in its broadest aspect, a multiple-ring communication system of the kind referred is characterised in that the receiver/transmitter equipments of each node are able to transmit on the rings ring test messages the destination of which is the node itself, the processing means controlling the receiver/transmitter equipments of the node and being able to check the receipt or absence of receipt of the ring test messages prior to possibly executing the reconfiguration operations.

By checking the receipt or absence of receipt of the ring test messages information is obtained on the condition of all the nodes and of all the ring portions interconnecting these nodes. Thus each node always knows if one or both rings are operating correctly or not and if loops have been established between these rings or not.

Another characteristic feature of the present system is that the receiver/transmitter equipments of each node are also able to transmit on the rings neighbouring node test messages which upon receipt by receiver/transmitter equipments of said neighbouring nodes normally give rise to the transmission to the node of node test reply messages on rings different from those on which the neighbouring node test messages were transmitted, the processing means being also able to check the receipt or absence of receipt of the test reply messages prior to possibly executing the reconfiguration operations.

By checking in each node the receipt or absence of receipt of the node test reply messages information is obtained therein on the condition of the receiver and transmitter ring portions connecting the transmitters and receivers of this node to the receivers and transmitters of the neighbouring nodes, of its own receivers and transmitters and of the receivers and transmitters of the neighbouring nodes.

With all this information available, well founded decisions can be taken by the processing means.

An embodiment of the invention will now be described by way of example with reference to the accompanying drawings wherein:

Fig. 1 shows a multiple-ring communication system according to the invention;

Fig. 2 represents a station or node N1 and associated user circuits G1 of Fig. 1 in more detail.

The multiple-ring communication system shown in Fig. 1 is more particularly a double-ring system which includes a plurality of stations or nodes N1 to Nn which are intercoupled so as to form two communication loops or rings RG1 and RG2 and which operate on an equal basis without central control. The system operates in a plesochronous way, meaning that each node is operating at the rhythm of its own clock but that the frequency difference between the clocks of all nodes is restricted to

a predetermined value. Such a way of operation is generally known in the technique e.g. in packet switching networks. On the rings RG1 and RG2, messages are able to be transmitted in opposite directions as indicated by the arrows and in each node a message receiver and a message transmitter are connected to each of the rings, e.g. in N1, R11 and T11 are connected to RG1 and R12 and T12 are connected to RG2. The nodes N1 to Nn have access to groups of user terminal circuits G1 to Gn respectively.

Node N1 represented in Fig. 2 is identical to all other nodes and includes equipment associated in common with both the rings RG1 and RG2 as well as equipment individually associated with each of these rings.

The common equipment of node N1 includes:

— a main processor MP1, e.g. 8086 INTEL processor, having access to an individually associated memory MEM1 including a system status register SSR and to a common communication bus BB1;

— 16 auxiliary processors AP100 to AP115 having access to the common bus BB1 and to individually associated sets of user terminal circuits S100 to S115 respectively. These sets together form the above mentioned group G1.

The equipment of node N1 individually associated with ring RG1 includes:

— a receiver R11 and a transmitter T11 both connected to RG1;

— a receiver processing unit RPU11 directly coupled to receiver R11 and a transmitter processing unit TPU11 coupled to transmitter T11 via a change-over switch SW11 which will be considered later;

— a delay buffer constituted by a first-in-first-out or FIFO circuit FIFO11 coupling RPU11 to TPU11;

— an input buffer or receiver memory RM11 and an output buffer or transmitter memory TM11 coupled on the one hand to RPU11 and TPU11 respectively and on the other hand to the common bus BB1;

— a control register CR11 coupled to the latter bus BB1 via control lead c111 and storing a so-called "faulty side bit" FSB11. This bit is 0 or 1 depending on N1 at the side looking in the direction of RG1 (Fig. 1) being not faulty or faulty respectively. This side includes portions of RG1 and RG2 and node N2.

Likewise, the equipment individually associated with ring RG2 includes a receiver R12, a transmitter T12, a receiver processing unit RPU12, a transmitter processing unit TPU12; a change-over switch SW12, a delay buffer or FIFO circuit FIFO 12; an input buffer or receiver memory RM12, an output buffer or transmitter memory TM12 and a control register CR12 coupled to bus BB1 via control leads c112 and c113 and storing not only a malicious side bit FSB12 but also a so-called transparency bit TPB12. All these circuits are interconnected in the same way as the corresponding circuits of the equipment individually associated with RG1. FSB12 is 0 or 1 depending on the side of N2 looking in the direction of RG2 (Fig. 1) being not faulty or faulty respectively. This side includes portions of RG1 and RG2 and node Nn. Transparency bit TPB12 is associated with ring RG2 i.e. with RPU12 and is 1 or 0 depending on N1 being transparent or not respectively with respect to RG2 as will be explained later.

Change-over switch SW11 enables either TPU11 or TPU 12 to be connected to T11 and change-over switch SW12 likewise enables either TPU11 or TPU12 to be connected to T12. Each of these switches SW11 and SW12 is in fact a well known 2-input-1-output digital multiplexer further having a select input. The select inputs of these multiplexers are controlled by FSB12 and FSB11 stored in the control registers CR12 and CR11 respectively. Depending on FSB12 being 0 or 1, the output of TPU11 or TPU12 is connected to T11 and depending on FSB11 being 0 or 1, the output of TPU12 or TPU11 is connected to T12. In other words when FSB12 (FSB11) is 0 or 1, receiver R12 (R11) is coupled to transmitter T12 (T11) or T11 (T12) respectively, thus establishing a loop between RG2 (RG1) and RG1 (RG2).

Functions of the main processor MP1 are, amongst others, to collect via the bus BB1 messages transmitted by RM11, RM12 and the auxiliary processors AP100—AP115, to process these messages, to form new messages and store them via bus BB1 in TM11 or TM12 for subsequent transmission to another node, to set or reset the bits FSB11, FSB12, TMB12 via BB1 and c111, c112, c113 respectively, to transmit messages to the auxiliary processors via BB1, and to perform time-outs.

The auxiliary processors AP100—AP115 are able, for instance, to process messages incoming from MP1 via BB1 and from the associated sets of user terminal circuits S100—S115. They also can transmit messages to MP1 and to S100—S115.

The receiver processing circuits RPU11 and RPU12 are able to execute relatively simple functions such as processing the messages received in R11, R12 on RG1, RG2 and storing them either in RM11, RM12 (thus removing them from the ring) or in FIFO11, FIFO12 (for transmission to another node) depending on the type of message. In this connection it should be noted that a message can only be removed from the ring by the originating node and by the destination node.

Also, the transmitter processing units TPU11 and TPU12 are able to execute relatively simple functions such as processing the messages stored in TM11, TM12, or in FIFO11, FIFO12 and operating T11, T12 or T12, T11, depending on the condition of FSB11, FSB12, in order to transmit these messages on RG1, RG2 or RG2, RG1.

The messages used in the present system for testing it are the following:

— ring test messages which are intended for testing the ring RG1 and RG2. They are transmitted in each node on RG1 and RG2 and removed from the ring when received back in this node. If everything is working properly this happens after the message has passed through all the other nodes via RG1 or

RG2. It is clear that when a ring test message is received back in a node this is an indication that the receiver/transmitter equipments, associated with this ring, of all the nodes operate correctly and that the portions of the ring interconnecting these nodes are not faulty;
— adjacent-node test messages which are intended mainly for testing the adjacent nodes of each node. They are transmitted in each node on RG1 or RG2;
— adjacent-node test reply messages which are also mainly intended for testing the adjacent nodes of a node as they are transmitted from these adjacent nodes to the latter node in reply to adjacent-node test messages. This transmission is performed on rings different from those on which the adjacent-node test messages were received by the node. Hence, when an adjacent node test reply message is received in a node from an adjacent node, this is an indication that the receiver/transmitter equipments involved in the operation of both nodes and the ring portions interconnecting these nodes are not faulty.

In what follows, the adjacent-node test and the adjacent-node test reply messages are called node test and node test reply messages for simplicity.

More particularly, the equipment of node N1 is able to transmit two ring test messages RTM11 (message from N1 on RG1) and RTM12 (message from N1 on RG2) and two node test messages NTM11 (message from N1 on RG1) and NTM12 (message from N1 on RG2) and to receive from the adjacent nodes N2 and Nn in reply to these node test messages two node test reply messages NTRM22 (message from N2 on RG2) and NTRMn1 (message from Nn on RG1). Obviously the node N1 is also able to transmit node test reply messages NTRM11 and NTRM12 to N2 and Nn respectively, but this is without importance for the further description.

The above test messages may contain:
— a message type indication;
— a source node address;
— a destination node address;
— a ring identifier for identifying the ring on which the message is transmitted;
— information related to the message type, e.g. an indication that upon the message being received on a ring a reply message should be transmitted on the other ring. If there are more than two rings the identity of this other ring should be given.

The ring test messages RTM 11 and RTM12 contain the following information:
— the message type indication: RTM;
— the source node address: N1;
— the destination node address: N1;
— ring identifier; RG1 and RG2 respectively.

The node test messages NTM11 and NTM12 contain the following information:
— message type indication: NTM;
— source node address: N1;
— ring identifier: RG1 and RG2 respectively;
— information for the main processor indicating that upon the receipt of the message by a node on a ring a reply message should be transmitted on the other ring to the source node of the message. The information can also contain an indication that upon the receipt of the message a test programme should be started and that a reply message should be transmitted to the source node of the test message on the other ring only when the test was successful. To be noted that the node test message does not contain a destination address, i.e. the address of an adjacent node. In this way, each node can be perfectly ignorant of the structure of the ring and more particularly has not to know the identity of its neighbouring nodes. Thus the structure of the ring can be modified without it being necessary to adapt the messages.

The node test reply messages NTRM22 and NTRMn1 contain the following information:
— the message type indication: NTRM;
— the source node address: N2 and Nn respectively;
— the destination address: N1;
— the ring identifier: RG2 and RG1 respectively.

In brief, the above system operates as follows:
— each message and hence each of the test messages RTM11, NTRMn1 (RTM12, NTRM22) received by R11 (R12) is processed by RPU11 (RPU12) and stored in RM11 (RM12) or in FIFO11 (FIFO12) depending on the message being intended for node N1 or for another node respectively;
— a received message which has been stored in RM11 (RM12) is processed by MP1 and possibly transmitted to an auxiliary processor AP100—AP115 where it is further processed and possibly sent from there to one of the user terminal circuits of the associated set S100—S115 respectively;
— a received message which has been stored in FIFO11 (FIFO12) is processed by TPU11 and transmitted on RG1 (RG2) by transmitter T11 (T12) if FSB12 (FSB11) is 0. However, if FSB12 (FSB11) is 1 this message is transmitted on RG2 (RG1) by transmitter T12 (T11).
— a message which has been locally formed by MP1 and has to be transmitted to another node, such as the test messages RTM11, RTM12, NTM11, NTM12, is stored by MP1 either in TM11 or in TM12 and then handled by TPU11 or TPU12 in a similar way as the messages stored in FIFO11 and FIFO12.

From the nature of the above test messages it follows that in normal circumstances:

—the ring test messages RTM11 and RTM12 test the following:
= all the portions of RG1 and RG2 interconnecting the various nodes;
= in N1 the two receiving/transmission equipments;
= in the other nodes N2 to Nn the equipment corresponding to the following equipment in N1: R11, RPU11, FIFO11, TPU11, SW11, T11 and R12, RPU12, FIFO12, TPU12, SW12, T12;
— the node test messages NTM11 and NTM12 test the following:
= the portion of RG1 interconnecting N1 and N2 and the portion of RG2 interconnecting N1 and Nn;
= in N1 the two transmission equipments;
= in N2 and Nn the equipments corresponding to the following receiving equipments in N1:R11, RPU11, RM11, MP1, MEM1; and R12, RPU12, RM12, MP1, MEM1 respectively.
— the node test reply messages NTRM22 and NTRMn1 test the following:
= the portion of RG2 interconnecting N1 and N2 and the portion of RG1 interconnecting N1 and Nn;
= in N2 and Nn the transmission equipment corresponding to the following transmission equipments in N1:
MP1, MEM1, TM12, TPU12, SW12, T12; and MP1, MEM1, TM11, TPU11, SW11, T11 respectively
= in N1 the two receiving equipments.

Returning to the main processor MP1, the latter is more particularly able to control the following functions:
— making an interconnection or loop between RG1 and RG2, i.e. between R11 and T12, by setting to 1 FSB11 which then controls SW12 in such a manner that TPU11 is connected to T12;
— making a loop between RG2 and RG1, i.e. between R12 and T11, by setting to 1 FSB12 which then controls SW11 in such a way that TPU12 is connected to T11;
— opening of a loop between RG1 and RG2 resetting FSB11 to 0;
— opening of a loop between RG2 and RG1 by resetting FSB12 to 0;
— making RG2 transparent by setting to 1 TPM12 which controls RPU12;
— removing transparency from RG2 by resetting TPB12 to 0;
— transmitting messages to a maintenance system (not shown), e.g. to indicate that RG1 or RG2 is working correctly or is faulty;
— updating system status register SSR shown in columns 2 to 5 of the table on the last page of the description. SSR indicates the status of the system as seen in node N1. More particularly it has four successive locations for registering the reaction detected in node N1 to the transmission of RTM11, RTM12, NTM11 and NTM12 respectively. This reaction is one of the following codes:
1) 00 or N: this means that no message has been received
2) 01 or RTM11, RTM12: this means that the ring test message RTM11, RTM12 has been received on ring RG1, RG2 respectively, i.e. by RM11, RM12;
3) 10 or RTM'11, RTM'12: this means that ring test message RTM11, RTM12 has been received on ring RG2, RG1 respectively, i.e. by RM12, RM11;
4) 11 or NTRM22, NTRMnm: this means that node test reply message RTM11, RTM12 has been received from N2, Nn respectively.

The codes 00 to 11 are not shown in the SSR.

Main processor MP1 alternately counts time intervals TO1 and TO2. During TO1 normal messages are transmitted and no tests are performed, whilst, during TO2, only ring test messages RTM11 and RTM12 and node test messages NTM11 and NTM12 are transmitted. The receipt or absence of receipt of the ring test messages as well as of the node test reply messages NTRM22 and NTRMn1 is checked and their possible receipt is registered in the system status register SSR. In normal circumstances, these test messages are received before the end of TO2 at which moment the processor MP1 evaluates the result of the tests by checking the contents of SSR. As a function of this result MP1 then decides to execute measures to reconfigure a faulty system back to a correctly operating system to inform a maintenance system (not shown), or to do nothing.

The operation of the above system is described in more detail hereinafter, it being supposed that, in node N1, bits FSB11, FSB12 and TPB12 are 0. This means that TPU11 and TPU12 are connected to T11 and T12 via SW11 and SW12 respectively and that ring RG2 in N1 is not transparent.

After the end of a time interval TO1, the main processor MP1:
— resets the system status register SSR by writing in each of the 4 locations thereof the code N;
— starts counting time interval TO2;
— forms test messages RTM11, NTM11 and RTM12, NTM12 and stores them in TM11 and TM12 respectively together with an order of transmission.

The transmitter processing units TPU11 and TPU12 then first read the ring test messages RTM11 and RTM12 and operate T11 and T12 to transmit these messages on RG1 and RG2 respectively. Immediately afterwards and in an analogous way the node test messages NTM11 and NTM12 are read and transmitted on RG1 and RG2 respectively.

The ring test message RTM11 on RG1 is successively received and transmitted by the nodes N2 to Nn and finally received by node N1, and the ring test message RTM12 on RG2 is successively received and transmitted by the nodes Nn to N2 and finally received by N1.

More particularly, when a ring test message RTM11, RTM12 is received by a receiver of one of the

nodes N2 to Nn the receiver processing unit associated with this receiver deduces from the contents of the message that its destination is node N1 and therefore writes the ring test message in the associated FIFO. The processing unit associated with this FIFO starts reading it as soon as e.g. a byte of this message has been stored in the FIFO and then operates the associated transmitter to transmit it on the associated ring. The following bytes of the message are transmitted in the same way. However, when a ring test message RTM11, RTM12 is received by a receiver of node N1 the receiver processing unit associated with this receiver deduces from the contents of the message that it is intended for node N1 and therefore stores it in its associated memory or input buffer. Thus the ring test message is removed from the ring.

The node test messages NTM11 and NTM12 transmitted on the rings RG1 and RG2 are received by the receivers associated with RG1 and RG2 of the neighbouring nodes N2 and Nn respectively. The receiver processing unit associated with each of these receivers deduces from the contents of the message received that it is a node test message and therefore stores it in the associated memory or input buffer. Thus the node test message is removed from the ring. When each of the main processors of the nodes N2 and Nn afterwards reads the node test message from this memory it forms a node test reply message NTRM22 and NTRMn1 respectively, and stores it in the transmitter memory associated with the ring, RG2 and RG1 respectively, for transmission on this ring. Possibly the transmission of a reply message can be dependent on the successful execution of tests in N2, Nn. Finally, the transmitter processing unit associated with this transmitter memory, after having read this message, operates the associated transmitter to transmit the message on the ring, RG2 or RG1 respectively. These node test reply messages NTRM22 and NTRMn1 are received by the receivers R12 and R11 of node N1 and then stored in RM12 and RM11 by RPU12 and RPU11 respectively. Thus they are removed from the rings RG1 and RG2.

Each time a ring test message or a node test reply message is received in one of the receiver memories RM11 and RM12, the main processor MP1 updates its system status register SSR. With regard to the ring test messages it should be noted that the processor MP1 upon receiving such a message knows if it is received on the ring on which it was originally transmitted or not since the ring on which it is received is that associated with the memory RM11, RM12 in which the received message is stored, whilst the ring on which it was transmitted is given by the ring identifier of the message.

If everything is correct the ring test messages RTM11 and RTM12 as well as the node test reply messages NTRM21 and NTRMn1 are received in node N1 before the end of the time interval TO2 on the rings RG1, RG2, RG2 and RG1 respectively. This is not so if one or more nodes and/or one or more ring portions between these nodes are faulty. The above mentioned table lists the various possible contents of the SSR which may be found in node N1 at the end of a time interval TO2 following the transmission of the ring messages RTM11 and RTM12 and of the node test messages NTM11 and NTM12. Hereby X means either N or NTRM22 and Y means either N or NTRMn1.

From the information stored in SSR the main processor MP1 draws the following conclusions with regard to the system status SS:

RG1OK, RG1NOK: ring RG1 is OK (correct), NOK (not correct) respectively;

RG2OK, RG2NOK: ring RG2 is OK, NOK respectively;

N2D, NnD: adjacent node N2, Nn is not operative or down, e.g. power down. This condition includes the condition of the node N2, Nn and/or the ring portions between N2, Nn and N1;

N2O, NnO: adjacent node N2, Nn is on-line, but not active, e.g. during programme-loading;

N2A, NnA: adjacent node N2, Nn is active. This also includes the node N2, Nn and the ring portions between N2, Nn and N1;

NRG1, NRG2: in a node able to be reached from node N1 via RG1, RG2 a loop has been established, between RG1 and RG2 or between RG2 and RG1 respectively.

To take the information of SSR into account the main processor MP1 possibly controls the execution of reconfiguration operations in node N1 by setting to 1 one or more of the bits FSB11, FSB12 and TPB12.

The various cases which may be found in SSR and which are shown in the above table are briefly considered hereinafter.

CASE 0

This is the rest condition where the contents of SSR are: N, N, N, N.

CASE 1

RTM11 and RTM12 are received back in N1 or RG1 and RG2 respectively. Independently of the receipt or absence of receipt of NTRM22 or NTRMn1, as indicated by X or Y, the main processor MP1 deduces from this information that in the system both the rings RG1 and RG2 are O.K. and decides therefore that no reconfiguration measures have to be taken.

It also sends a message to the maintenance system to inform it of the fact that both the rings are O.K.

CASES 2 AND 3

Either only RTM11 or RTM12 is received back in N1 on RG1 and RG2 respectively. Again

independently from the receipt or absence of receipt of the test reply messages, MP1 deduces from this information that in the system either ring RG1 or RG2 is O.K. and decides therefore that no reconfiguration measures have to be taken. It also informs the maintenance system of the fact that ring RG2 or RG1 is not O.K.

CASE 4

RTM11 and RTM12 are received back in N1 on RG2 and RG1 as indicated by RTM'11 and RTM'12 respectively. Independently of the reception or absence of reception of NTRM22 or NTRMn1, as indicated by X or Y, the main processor MP1 deduces from this information that RG1 and RG2 are both interrupted somewhere, e.g. in Z on Fig. 1, but that, as indicated by NRG1 and NRG2 in nodes Nx and Nn-1 (Fig. 1) able to be reached from N1 via RG1 and RG2, loops Lx and Ln-1 have been established between RG1 and RG2 and between RG2 and RG1 respectively so that a new single ring has been formed. Consequently, MP1 then only transmits test messages and sets to 1 the transparency bit TPB12 which makes RG2 in Nn transparent by bringing RPU12 in such a condition that all messages, except node test reply messages, received by R12 are stored in FIFO12 for further transmission by T12 or RG2. Node test reply messages are stored in RM12. Making RG2 transparent in N1 in the case where two loops, such as Lx and Ln-1, have been formed is necessary because — as mentioned above — messages may be removed from a ring either by the node which originated the message or by the node of destination. Indeed, suppose that N1 transmits a message on RG1 to Nn-1 then this message will be looped back by Lx to N1 and removed from the ring if no precautions are taken. However, as RG2 is transparent in node N1 this message is immediately transmitted by T12 to Nn and from there to the destination node Nn-1.

CASES 5 TO 8

RTM11 is received back in N1 on RG2, as indicated by RTM'11, and RTM12 is not received back in N1, NTRM22 and NTRMn1 are either received or not.

From this information the main processor MP1 deduces the system status SS as follows:

— from the receipt of RTM11 on RG2 it follows that RG1 is interrupted somewhere but that in some node able to be reached via RG1 a loop has been established between RG1 and RG2. This is indicated by NRG1;

— from the absence of receipt of RTM12 it follows that RG2 is interrupted somewhere;

— from the receipt of NTRM22 or NTRMn1 it follows that N2 or Nn is active. This is indicated by N2A or NnA;

— from the absence of receipt of NTRM22 or NTRMn1 it follows that N2 or Nn is down. This is indicated by N2D or Nnd;

— from the above it follows that

= in case 5 the side of N1 looking in the direction of RG2 is faulty;

= in cases 6 and 8 both the sides of N1 are active;

=in case 7 both the sides of N1 are not active.

Based on the above-detected system status, MP1 decides to set to 1 both FSB11 and FSB12 for the cases 5 and 7 and to do nothing for the cases 6 and 8. By setting FSB11 to 1 a loop is established between R11 and T12, i.e. between RG1 and RG2, whilst, by setting FSB12 to 1, a loop is established between R12 and T11, i.e. between RG2 and RG1. The latter loop is required to permit messages received from N2 by R12 to be transmitted back to N2, whilst the former loop enables messages sent by Nn after it is no longer down to be transmitted to N1 and received back therefrom and to prevent such messages from being transmitted by T11.

It should be further noted that cases 6 and 8 are unstable cases because as RG2 is interrupted somewhere a loop will be formed between RG2 and RG1 in which case also RTM12 will be received on another ring as RTM'12 (see case 4).

CASES 9 TO 12

These cases are similar to cases 4 to 8, the roles of N2 and Nn being, however, interchanged. For this reason the system status for the cases 9 to 12 corresponds to the system status for the cases 8, 6, 7, 5 respectively and the same is true for the reconfiguration measures.

CASES 13 TO 16

RTM11 and RTM12 are not received back in N1 and NTRM22 and NTRMn1 are either received or not so that the system status for the cases 13 to 16 is similar to that of cases 5 to 8 respectively, except for NRG1 which now does not form part of the status and for N20 which is replaced by N2D. Also the reconfiguration measures for the cases 13 to 15 are identical to those for the cases 5 to 7. However for case 16, contrary to case 8, both FSB11 and FSB12 are now also set to 1. No messages are sent to the maintenance system.

CASES 17 AND 18

These cases are unstable cases. Indeed, although there exists a correct ring RG1 (case 17) or RG2 (case

18) as indicated by RTM11 or RTM12 respectively, there still exists a loop as indicated by RTM'11 or RTM'12 which has to be opened as it is of no use. This loop was established during the time both rings were faulty.

O	SSR				SS	Reconfiguration R12–T11	Reconfiguration R11–T12	Maintenance messages and/or comments
	RTM11 N	RTM12 N	NTM11 N	NTM12 N				
1	RTM11	RTM12	X	Y	RG1OK, RG2OK	–	–	RG1OK RG2OK
2	RTM11	N	X	Y	RG1OK	–	–	RG2NOK
3	N	RTM12	X	Y	RG2OK	–	–	RG1NOK
4	RTM'11	RTM'12	X	Y	NRG1, NRG2 N2O of N2A, NnO of NnA	–	TPB12=1	Single ring with transparency
5	RTM'11	N	NTRM22	N	NRG1, N2A, NnD	MSB11=1	MSB12=1	–
6	RTM'11	N	NTRM22	NTRMn1	NRG1, N2A, NnA	–	–	Unstable state
7	RTM'11	N	N	N	NRG1, N2O, NnD	MSB11=1	MSB12=1	–
8	RTM'11	N	N	NTRMn1	NRG1, N2O, NnA	–	–	Unstable state
9	N	RTM'12	NTRM22	N	NRG2, N2A, NnO	–	–	–
10	N	RTM'12	NTRM22	NTRMn1	NRG2, N2A, NnA	–	–	–
11	N	RTM'12	N	N	NRG2, N2D, NnO	MSB11=1	MSB12=1	–
12	N	RTM'12	N	NTRMn1	NRG2, N2O, NnA	MSB11=1	MSB12=1	–
13	N	N	NTRM22	N	N2A, NnD	MSB11=1	MSB12=1	–
14	N	N	NTRM22	NTRMn1	N2A, NnA	–	–	–
15	N	N	N	N	N2D, NnD	MSB11=1	MSB12=1	–
16	N	N	N	NTRMn1	N2D, NnA	MSB11=1	MSB12=1	–
17	RTM11	RTM'12	X	Y	RG1OK, NRG2	–	–	Unstable state
18	RTM'11	RTM12	X	Y	NRG1, RG2OK	–	–	Unstable state

GB 2 133 952 A

CLAIMS

1. A multiple-ring communication system of the kind which includes a plurality of nodes intercoupled in at least two rings and operating on an equal basis without central control, each node including a plurality of message receiver/transmitter equipments associated with distinct ones of the rings respectively and able to receive messages on the rings and transmit messages thereon in opposite directions, and processing means for checking the receipt or absence of receipt of signals on the receiver ring portions connecting the receivers of the node to transmitters of its neighbouring nodes and for performing reconfiguration operations in response to the result provided by the checking operation, characterised in that the receiver/transmitter equipments of each node (N1) are able to transmit on the rings (RG1, RG2) ring test message (RTM11, RTM12) the destination of which is the node (N1) itself, the processing means (MP1, MEM1) controlling the receiver/transmitter equipments of the node and being able to check the receipt or absence of receipt of the ring test messages prior to possibly executing the reconfiguration operations.

2. A multiple-ring communication system according to claim 1, characterised in that the receiver/transmitter equipment of each node (N1) are also able to transmit on the rings (RG1, RG2) neighbouring node test messages (NTM11, NTM12) which upon receipt by receiver/transmitter equipments of the neighbouring nodes (N2, Nn) normally give rise to the transmission to the node of node test reply messages (NTRM22, NTRMn1) on rings different from those on which the neighbouring node test messages were transmitted, the processing means (MP1, MEM1) being also able to check the receipt or absence of receipt of the test reply messages prior to possibly executing the reconfiguration operations.

3. A multiple-ring communication system according to claim 1, characterised in that each of the message receiver/transmitter equipments (R11, RPU11, RM11, FIFO11, TPU11, SW11, T11, TM11; R12, RPU12, RM12, FIFO12, TPU12, SW12, T12, TM112) includes receiver means (R11, RPU11,; R12, RPU12), transmitter means (TPU11, T11, SW11,; TPU12, T12, SW12), a delay buffer (FIFO11; FIFO12), an input buffer (RM11; RM12) and an output buffer (TM11; TM12), the receiver means having an input coupled to a respective ring (RG1, RG2) and outputs coupled to inputs of the input and delay buffers and the delay and output buffers having access to the transmitter means, the output buffer having an input coupled to a bus (BB1) and the transmitter means having an output having access to the ring, and the processing means (MP1, MEM1) being coupled to the bus which is common to both the receiver/transmitter equipments.

4. A multiple-ring communication system according to claim 3, characterised in that the transmitter means (TPU11, T11, SW11; TPU12, T12, SW12) of each of the receiver/transmitter equipments includes a transmitter circuit (T11, T12), an associated transmitter processing circuit (TPU11, TPU12) and an associated switching means (SW11; SW12) controlled by the processing means (MP, MEM1), the delay (FIFO11; FIFO12) and output (TM11; TM12) buffers being coupled to the transmitter processing circuits and each of the transmitter processing circuits (TPU11; TPU12) being coupled to both transmitter circuits (T11, T12) via the associated switching means (SW11, SW12), and the processing means (MP1, MEM1) controlling the switching means (SW11, SW12) in such a way that each of the transmitter processing circuits (TPU11, TPU12) can be effectively connected to either one of the transmitter circuits (T11, T12).

5. A multiple-ring communication system according to claim 4, characterised in that each of the switching means (SW11; SW12) is a two-input-one-output digital multiplexer with two inputs coupled to the outputs of the transmitter processing circuits (TPU11, TPU12), with an output coupled to the input of a distinct one of the transmitter circuits (T11, T12), and with a select input controlled by a first bit (FSB12; FSB11) able to be activated or de-activated by the processing means (MP1; MEM1) in order to connect effectively the one or the other of the two inputs to the one output and to establish a loop or not between the rings.

6. A multiple-ring communication system according to claim 3, characterised in that the receiver means (R11, RPU11,; R12, RPU12) of each of the receiver/transmitter equipments includes a receiver circuit (R11, R12) and an associated receiver processing circuit (RPU11; RPU12), at least one (RPU12) of the receiver processing circuits (RPU11; RPU12) being able to be brought into such a condition by the processing means that it is only able to store received node test reply messages (NTRM22) in its associated input buffer (RM12).

7. A multiple-ring communication system according to claim 6, characterised in that the one receiver processing circuit (RPU12) is controlled by a second bit (TPB12) able to be activated by the processing means (MP1, MEM1) to bring that one receiver processing circuit in the condition wherein it is only able to store received node test reply messages in its associated input buffer.

8. A multiple-ring communication system according to claim 7, characterised in that when the second bit (TPB12) has been activated by the processing means (MP1, MEM1) the latter means still control the transmission of the ring test and node test messages.

9. A multiple-ring communication system according to claim 1, characterised in that each of the test messages includes at least an indication of the type of message (RTM, NTM, NTRM), a source node address indicating the node originating the message, and a ring identifier identifying the ring on which the message is transmitted.

10. A multiple-ring communication system according to claim 9, characterised in that the ring test message also includes the address of the node of destination, this address being equal to the source node address.

11. A multiple-ring communication system according to claim 9, characterised in that the neighbouring node test message includes an indication that a node test reply message has to be transmitted on a predetermined ring, but does not include the address of a node of destination.

12. A multiple-ring communication system according to claim 11, characterised in that the neighbouring node test message includes an indication that a test has to be performed in the node and that a node test reply message has only to be transmitted when this test has been successful.

13. A multiple-ring communication system according to claim 9, characterised in that in each node (N1) the processing means (MP1, MEM1) are able to control regularly (TO2) the transmission by the receiver/transmitter/equipments of the ring test messages (RTM11, RTM12) and of the neighbouring nodes test messages (NTM11, NTM12), to store in a register (SSR) the status of the system by storing therein the receipt of the ring test messages and of the neighbouring nodes test reply messages (NTRM22, NTRMn1) as well as the fact that the ring test messages are received on the same ring or on a different ring from that on which they were transmitted, the fact being derived from the identity of the input buffer (RM11, RM12) in which a ring test message is stored by the receiver means (R11, RPU11; R12; RPU12) and from the ring identifier of the ring test message, and to (TO2) evaluate regularly the system status from the thus registered status information and perform the reconfiguration operations (FSB11, FSB12, TPB12) in function of the system status.

14. A multiple-ring communication system according to claim 13, characterised in that one of the reconfiguration operations consists in activating at least one of the first bits (FSB11, FSB12) so as to establish a loop between the rings.

15. A multiple-ring communication system according to claim 14, characterised in that both the first bits (FSB11, FSB12) are activated when the registered status information (SSR, cases 5, 7, 11, 12) indicates that a ring test message transmitted on a ring has been received on another ring, and that no ring test message and no neighbouring node test reply message have been received on the one ring.

16. A multiple-ring communication system according to claim 14, characterised in that the first bits (FSB11, FSB12) are both activated when the registered status information (SSR, cases 13, 15, 16) indicates that no ring test messages and only one neighbouring node test reply message have been received.

17. A multiple-ring communication system according to claim 1, characterised in that the receiver means (RPU11, R11; RPU12, R12) of a node (N1) are able to store a received message in the input buffer (RM11, RM12) when the address of the source node or of the destination node corresponds to that of the node (N1).

18. A multiple-ring communication system according to claim 17, characterised in that another one of the reconfiguration operations consists in activating the second bit (TPB12) when the registered status information (SSR, case 4) indicates that both ring test messages have been received on different rings from those on which they were transmitted.

19. A multiple-ring communication system substantially as described with reference to the accompanying drawings.

Printed for Her Majesty's Stationery Office by the Courier Press, Leamington Spa, 1984. Published by the Patent Office, 25 Southampton Buildings, London, WC2A 1AY, from which copies may be obtained.